Bayerische Julius-Maximilians-Universität
Würzburg
Institut für Informatik
Lehrstuhl für Verteilte Systeme
Prof. Dr. P. Tran-Gia

Wireless LAN Performance Studies in the Context of 4G Networks

Klaus Heck

Würzburger Beiträge zur
Leistungsbewertung Verteilter Systeme
Bericht

Würzburger Beiträge zur Leistungsbewertung Verteilter Systeme

Herausgeber

Prof. Dr. P. Tran-Gia
Universität Würzburg
Institut für Informatik
Lehrstuhl für Verteilte Systeme
Am Hubland
D-97074 Würzburg

Tel.: +49-931-888-6630
Fax.: +49-931-888-6632
email: trangia@informatik.uni-wuerzburg.de

Satz

Reproduktionsfähige Vorlage vom Autor.
Gesetzt in LaTeX Computer Modern 9pt.

ISSN 1432-8801

Wireless LAN Performance Studies in the Context of 4G Networks

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Bayerischen Julius–Maximilians–Universität Würzburg

vorgelegt von

Klaus Heck

aus

Gerolzhofen

Würzburg 2005

Eingereicht am: 28. Januar 2005
bei der Fakultät für Mathematik und Informatik
1. Gutachter: Prof. Dr.-Ing. P. Tran-Gia
2. Gutachter: Prof. Dr. rer. nat. habil. Carmelita Görg
Tag der mündlichen Prüfung: 13. Juli 2005

Danksagung

Obwohl diese Arbeit ausschließlich meinen Namen trägt, ist für ihre Entstehung die Hilfe vieler anderer Personen nötig gewesen. Diese sollen natürlich nicht ungenannt bleiben.

Zuallererst möchte ich im Rahmen dieser Danksagung meinen Doktorvater Prof. Dr.-Ing. Phuoc Tran-Gia erwähnen. Seine Unterstützung und seine reichlichen Kontakte zu Wissenschaft und Industrie haben es mir erlaubt in einem Gebiet zu forschen, das in den Forschungs- und Entwicklungsabteilungen der Industrie immer mehr an Bedeutung gewinnt. Die Idee entstand, wie wohl sehr häufig, bei einer ausgiebigen, abendlichen Diskussion im Nachtleben Berlins mit einem sehr guten Bekannten meines Doktorvaters, Dr. Nikhil Jain.

Herr Tran-Gia hat es aber auch geschafft an unserem Lehrstuhl ein Arbeitsumfeld zu schaffen, das von kollegialer Stimmung und kooperativer Atmosphäre geprägt ist. Kollegen, Diplomanden, Studenten und Praktikanten pflegen einen freundschaftlichen, aber dennoch professionellen, Umgang. So enden viele Schwätzchen im Gang oder beim Espresso doch wieder bei einer wissenschaftlichen Diskussion, die oft verschiedene Forschungsgebiete am Lehrstuhl umfasst und so den Horizont eines jeden einzelnen erweitern hilft.

Dabei gebührt meinen Kollegen ein besonderer Dank: Dr. Kenji Leibnitz, Dr. Dirk Staehle, Andreas Maeder, Tobias Hoßfeld, Rastin Pries, Dr. Michael Menth, Jens Milbrandt, Andreas Binzenhöfer, Robert Henjes, Rüdiger Martin, Simon Oechsner, Bartosz Wagner, Prof. Dr. Oliver Rose, Dr. Kurt Tutschku, Dr. Norbert Vicari, Dr. Mathias Dümmler und Stefan Köhler. Neben den fachspezifischen Aktivitäten waren die privaten Treffen immer ein besonderes Erlebnis.

Für die unbürokratische Unterstützung bei allen organisatorischen Angelegenheiten möchte ich mich bei Frau Gisela Alt bedanken. Damit hat sie mir und meinen Kollegen doch immer wieder sehr viel Zeit und vor allem Nerven erspart.

Einen wesentlichen Anteil an der Entstehung meiner Arbeit trugen auch alle studentischen Mitarbeiter, Praktikanten und Diplomanden bei. Ohne ihren Einsatz wäre es nicht möglich gewesen eine so umfangreiche Untersuchung anzufertigen. Deshalb geht mein herzlicher Dank an: Rastin Pries, Tom Wirth, Matthias Wiesen, Alexander Klein, Armin Lediger und Thomas Obeth.

Ebenso wichtig für den langen Weg eines Studiums mit anschließender Promotion ist das private Umfeld. Aus diesem Grund möchte ich meinen aufrichtigen Dank an meine Eltern Erika und Bruno Heck richten. Sie haben es mir erlaubt immer und ungefragt die nötige finanzielle und moralische Unterstützung zu bekommen.

Die Arbeit am Lehrstuhl ist meist zeitaufwändig und geprägt von Dienstreisen, und hinterläßt somit sicherlich auch seine Spuren im Alltagsleben. Der Verzicht auf einen Teil der üblichen Freizeit läßt sich da kaum verhindern. Deshalb gilt mein tiefster Dank meiner Partnerin Tomke Burger. Sie hat mich stets unterstützt und war mir damit in den vergangenen Jahren die wichtigste private Stütze.

Würzburg, im August 2005 *Klaus Heck*

Contents

Bayerische Julius-Maximilians-Universität
Würzburg

Institut für Informatik
Lehrstuhl für Verteilte Systeme
Prof. Dr. P. Tran-Gia

Wireless LAN Performance Studies in the Context of 4G Networks

Klaus Heck

Würzburger Beiträge zur
Leistungsbewertung Verteilter Systeme

Bericht

Würzburger Beiträge zur

Leistungsbewertung Verteilter Systeme

Herausgeber

Prof. Dr. P. Tran-Gia
Universität Würzburg
Institut für Informatik
Lehrstuhl für Verteilte Systeme
Am Hubland
D-97074 Würzburg

Tel.: +49-931-888-6630
Fax.: +49-931-888-6632
email: trangia@informatik.uni-wuerzburg.de

Satz

Reproduktionsfähige Vorlage vom Autor.
Gesetzt in LaTeX Computer Modern 9pt.

ISSN 1432-8801

Wireless LAN Performance Studies in the Context of 4G Networks

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Bayerischen Julius–Maximilians–Universität Würzburg

vorgelegt von

Klaus Heck

aus

Gerolzhofen

Würzburg 2005

Eingereicht am: 28. Januar 2005
bei der Fakultät für Mathematik und Informatik
1. Gutachter: Prof. Dr.-Ing. P. Tran-Gia
2. Gutachter: Prof. Dr. rer. nat. habil. Carmelita Görg
Tag der mündlichen Prüfung: 13. Juli 2005

1 Introduction

> The wireless telegraph is not difficult to understand. The ordinary telegraph is like a very long cat. You pull the tail in New York, and it meows in Los Angeles. The wireless is the same, only without the cat. Albert Einstein (1879-1955)

Wireless communication is nothing new. The first data transmissions based on electromagnetic waves have been successfully performed at the end of the 19th century. However, it took almost another century until the technology was ripe for mass market.

The first mobile communication systems based on the transmission of digital data were introduced in the late 1980s. Within just a couple of years they have caused a revolution in the way people communicate. The number of cellular phones started to outnumber the fixed telephone lines in many countries and is still rising. New technologies in 3G systems, such as UMTS, allow higher data rates and support various kinds of multimedia services.

Nevertheless, the end of the road in wireless communication is far from being reached. In the near future, the Internet and cellular phone systems are expected to be integrated to a new form of wireless system. Bandwidth requirements for a rich set of wireless services, e.g. video telephony, video streaming, online gaming, will be easily met. The transmission of voice data will just be another IP based service.

On the other hand, building such a system is by far not an easy task. The problems in the development of the UMTS system showed the high complexity of wireless systems with support for bandwidth-hungry, IP-based services. But the technological challenges are just one difficulty. Telecommunication systems are planned on a world-wide basis, such that standard bodies, governments, institutions, hardware vendors, and service providers have to find agreements and compromises on a number of different topics.

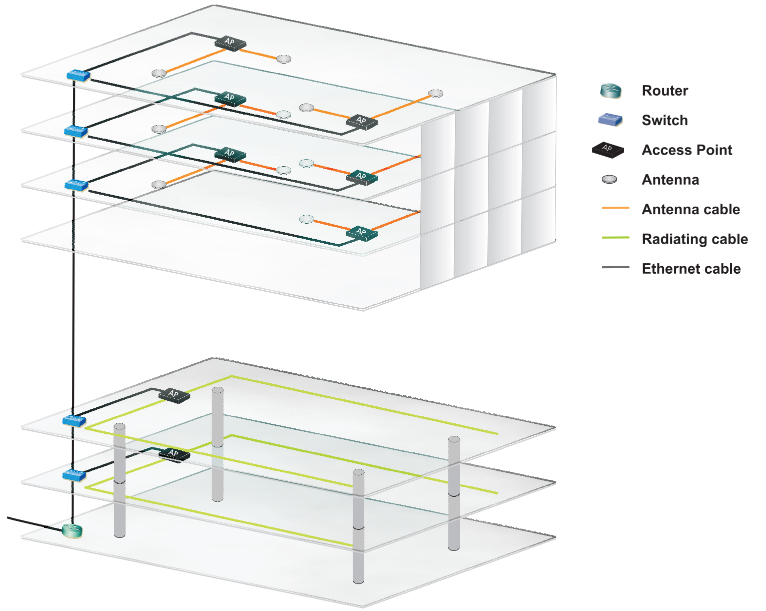

Figure 1.1: *Wireless LAN deployment in a large office building*

In this work, we provide the reader with a discussion of many of the topics involved in the planning of a Wireless LAN system that is capable of being integrated into the 4th generation mobile networks (4G) that is being discussed nowadays. Therefore, it has to be able to cope with interactive voice and video traffic while still offering high data rates for best effort traffic.

Let us assume a scenario as shown in Figure 1.1. A huge office complex is completely covered with Wireless LAN access points. Different antenna systems are applied in order to reduce the number of access points that are needed on the one hand, while optimizing the coverage on the other. No additional infrastructure is implemented. Our goal is to evaluate whether the Wireless LAN technology is capable of dealing with the various demands of such a scenario.

First, each single access point has to be capable of supporting best-effort and Quality of Service (QoS) demanding applications simultaneously. The IT infrastructure in our scenario consists solely of Wireless LAN, such that it has to allow users surfing the Web, while others are involved in voice calls or video conferences. Then, there is the problem of overlapping cells. Users attached to one access point produce interference for others. However, the QoS support has to be maintained, which is not an easy task. Finally, there are nomadic users, which roam from one Wireless LAN cell to another even during a voice call. There are mechanisms in the standard that allow for mobility, but their capabilities for QoS support are yet to be studied.

This shows the large number of unresolved issues when it comes to Wireless LAN in the context of 4G networks. In this work we want to tackle some of the problems. The remainder of this work is structured as follows.

Chapter 2 gives a short introduction to the history of wireless communication. The evolution from analog systems to new digital communication networks of the 3rd generation (3G) is summarized. This is followed by an overview of the different Wireless LAN standards and standard supplements. The specification of the IEEE 802.11 standards family is still an ongoing process. However, some interesting new approaches reached a rather stable state and their capabilities are evaluated in later chapters.

As explained above, a whole set of different applications has to be supported simultaneously. There is the traditional best-effort service for users surfing the Web or performing an FTP download. At the same time, other users run QoS demanding applications, such as voice or video transmissions. The different types of traffic that the system has to deal with are discussed in detail in Chapter 3. The Wireless LAN technology has to assure a certain level of satisfaction for all the users. However, quality assessment differs for each single traffic type. Chapter 3 explains the different approaches for quality assessment and summarizes the methods that are used in the remainder.

The Physical layer of the Wireless LAN technology is the topic of Chapter 4. Several modulation techniques that are defined for the two separate frequency bands of 2.4 GHz and 5 GHz are explained in detail. They have a major impact on the data rates that can be achieved. However, other properties of the wireless channel are of similar importance. All of these issues are discussed as well. The chapter is concluded by an overview of the simulation settings that are used later in order to study the performance of Wireless LAN.

The main contribution of this work to the discussion of Wireless LAN as a 4G technology is presented in Chapter 5. Different Medium Access Control (MAC) protocols that are defined for Wireless LAN are discussed and analyzed regarding their capability to support QoS demanding applications in large-scale environments. This includes cases where different applications are to be performed simultaneously in overlapping or co-located cells, as they definitely appear in the future. Later in this chapter, the impact of nomadic users on the performance capabilities is studied. This chapter finishes with the conclusion that the future Wireless LAN MAC protocol is capable of providing the necessary functionality, if it is properly configured.

However, the Wireless LAN protocol only deals with the ISO/OSI layer two. Nomadic users, on the other hand, might roam farther away from their home network, such that a handover on ISO/OSI layer three becomes necessary. The most important representative of the protocols supporting such IP handovers is Mobile IP. It is discussed in Chapter 6. But Mobile IP has been defined without QoS in mind. Therefore, it is not capable of supporting real-time applications. A large number of extensions to the basic approach have been published. Due to space limitation, it is not possible to present all of them, but the most important proposals are discussed as well.

Chapter 7 summarizes the work. It uses the results that were found in the different chapters and draws the final conclusion that Wireless LAN is an interesting technology and that it has the potential of being integrated into future mobile networks. Some issues are still considered open, but in terms of performance and service differentiation, the Wireless LAN Physical and MAC layers are ready for the future.

2 Short History of Wireless Communication

> It would appear that we have reached the limits of what is possible to achieve with computer technology, although one should be careful with such statements, as they tend to sound pretty silly in 5 years. John Von Neumann (1903 - 1957)

Computers started their revolution in the late 1930s when Konrad Zuse developed the Z1, the first computer the world has ever seen. It was little known outside of Germany, such that it had little impact on future computer development. However, transmission of data on the medium air was already known for more than 30 years by that time. Guglielmo Marconi presented his wireless telegraph in February 1896 to British telegraph authorities. But it was not until the computer technology reached a certain level of sophistication decades later, that the wireless technology could hit the mass market. In this chapter we give a short overview of the history of wireless communication. Note that this overview is by no means complete, but only mentions a few milestones. The last section in this chapter then gives an overview of the Wireless LAN standardization.

2.1 Analog Systems

The mobile phone systems of the first generation (1G) were all based on analog technologies ([Rap96], [Gib97]). In the beginning these systems could not be called cellular systems. Even though they consisted of several cell sites, seamless handovers could not be supported. The german "A-Netz" relied on manual switching. It was operated until 1977 and supported up to ten thousand users.

In the early seventies, the "B-Netz" was introduced. It was also based on frequency modulation and analog transmission, but it enhanced the "A-Netz" in terms of the switching technology. Now, no manual interaction was necessary any more as the system performed the switching automatically. However, a calling subscriber had to now the area code of the called customer's current location in order to initiate the connection. This type of network was operated until 1994.

The first real cellular networks were introduced in the early eighties. In 1981, the Siemens C450 standard was introduced in the german "C-Netz" . It was the first network that supported automatic handovers between the base stations. Although the transmission of speech was still performed analogously, the signaling information was already transmitted digitally. The C450 was restricted to Germany and Portugal.

In the United States, the Advanced Mobile Phone Service (AMPS) was introduced by AT&T in 1983. It was standardized in EIA/TIA-553. On the wireless interface it performs Frequency Division Multiple Access (FDMA) and frequency modulation in the band of 800 to 900 MHz with 30 kHz wide sub-bands (channels). Later the AMPS system was enhanced by the Narrowband Advanced Mobile Phone Service (NAMPS) as defined in IS-91. These two systems were mainly used in the United States.

Similar first generation analog cellular systems were also introduced in other parts of the world. This includes the Total Access Communication System (TACS) in the United Kingdom, Italy, Spain, Austria and Ireland, the Nordic Mobile Telephone (NMT) in the Scandinavian countries, the Radiocom 2000 in France, as well as the Nippon Telephone and Telegraph (NTT) or the JTACS/NTACS in Japan.

2.2 Digital Systems

After the analog systems as described above were operated for a couple of years, it became apparent that their capacity is simply too low to provide an adequate service to the increasing number of users. Especially in Europe a number of different incompatible analog cellular systems existed which made the interoperability an impossible task. Therefore, a common mobile communication standard throughout Europe was designed.

The European Telecommunications Standards Institute (ETSI) released phase 1 of the Global System for Mobile Communications (GSM) standard in 1990. It still forms the basis of the currently implemented systems. A variety of different services like telephony, emergency calls, conference calls, fax transmission, short messages, and data transmission at various rates up to 9600 bps could be offered. The standard clearly defines the different functional entities of the network, which leaves space for future enhancements ([Sch03]).

The GSM air interface implements Frequency Division Duplex (FDD) and Time Division Multiple Access (TDMA) scheme with up to eight simultaneous users sharing a single channel. Gaussian Minimum Shift Keying (GMSK) is used as the modulation scheme. In order to better overcome the problem of the adverse propagation conditions, slow frequency hopping is implemented. It assures, that long-term disruptions of a single frequency do not lead to dropped calls, but only to short-time packet losses which are hardly noticed by the users.

Initially, GSM utilized two 25 MHz frequency bands with a radio channel spacing of 200 kHz in the 900 MHz frequency band. Later, the 1800 MHz frequency band was added to the standard as the Digital Cellular System - 1800 (DCS1800). Even though the GSM standard was defined for Europe, it became a candidate for the U.S. Personal Communication Services (PCS) in the 1900 MHz band. Therefore, GSM managed to become the most important cellular phone system world-wide.

A similar development could be seen in the United States. The capacity of the AMPS systems was rapidly reaching its limits, such that a second generation (2G) digital system had to be defined for North America. This led to the development of the Interim Standard 54 (IS-54). It defines dual mode (AMPS/IS-54) mobile stations and base stations, thus, connecting the analog and digital worlds. The IS-54 standard specifies the utilization of TDMA and FDD technology with three simultaneous users per channel. Differential Quadrature Phase Shift Keying (DQPSK) modulation is used.

A number of additional standards were adopted to supplement the IS-54 definitions. The IS-41 standard deals with automatic roaming, intersystem handover, and other administrative tasks. IS-52 specifies the numbering plan, IS-53 defines supplementary services, and IS-93 defines interfaces to other systems. Nevertheless, the IS-54 standard still utilizes analog control channels in spite of the GSM systems. However, digital control channels exhibit a number of opportunities. They increase the capabilities for residential and in-building coverage or dramatically increase battery standby time. Therefore, the Electronic Industries Association / Telecommunications Industry Association (EIA/TIA) released the IS-136 interim standard for both the cellular (850 MHz) and the PCS (1900 MHz) frequency bands, to keep up with GSM capabilities. This system is also referred to as Digital Advanced Mobile Phone Service (D-AMPS).

The next step in the evolution of cellular systems began with the introduction of the Code Division Multiple Access (CDMA) technology by Qualcomm in 1989. At first, the industry was very skeptical regarding this new type of wireless access mechanism. However, in 1993 the Telecommunications Industry Association (TIA) adopted the cellular standard IS-95 based on CDMA. Compared to other technologies CDMA differs by its use of spread spectrum techniques for transmitting voice or data over the air. Rather than dividing RF spectrum into separate user channels by frequency slices or time slots, spread spectrum technology separates users by assigning them digital codes within the same broad spectrum. Advantages of CDMA technology include high user capacity and immunity from interference by other signals. Like TDMA IS-136, CDMA operates in the 1900 MHz band as well as the 800 MHz band. This basic CDMA approach is still referred to as 2G technology.

The first real beyond 2G technology was the General Packet Radio Service (GPRS). It defines an extension to the GSM system to support packet switching. GSM and GPRS could be supported by the same air interface; only the backbone network had to be extended in order to support circuit and packet switching simultaneously. Therefore, it was now possible to charge the data services on a volume basis rather than on a time basis. The GPRS enhanced networks are referred to as 2.5G technology.

Based on the IS-95 standard, also known as cdmaOne, the cdma2000 release 0 standard was published by "The Third Generation Partnership Project 2" (3GPP2) in August 1999. It defines the first real 3G network, based on CDMA technology and offers packet services with high data rates. The first commercial cdma2000 network was launched in Korea as early as October 2000. Since then, many other operators started their networks in a number of different countries all around the world. New variants of the cdma2000 standard have been developed in the meantime. The most important are cdma2000 1X, 1X EV-DV, 1X EV-DO, and cdma2000 3X. They all enhance the basic cdma2000 network in order to provide higher data rates while utilizing less spectrum.

The 3GPP2 project was mainly initiated by the International Telecommunication Union (ITU) as the counterpart to "The Third Generation Partnership Project" (3GPP) that was started by the European Telecommunications Standards Institute (ETSI) in 1998. The main goal of the 3GPP was *to produce globally applicable Technical Specifications and Technical Reports for a 3rd Generation Mobile System based on evolved GSM core networks and the radio access technologies that they support.* As early as in 1999, the first standard was finished. It is called the Universal Mobile Telecommunication System (UMTS) release 1999. It uses Wideband CDMA (WCDMA) as the air interface.

The first commercial UMTS networks were introduced by NTT DoCoMo in Japan in 2001. NTT refers to the technology as FOMA (Freedom Of Mobile multimedia Access). The chances for the introduction of new mobile services depending on high data rates seemed very prosperous. Therefore, many companies invested millions of dollars in order to buy the licenses needed for the frequency bands in all different countries. However, the dot com crash hit the market. Mobile business did not seem as prosperous any longer and companies greatly delayed the introduction of their UMTS services.

In parallel to the development of the cellular systems, a great deal of new developments were introduced in the area of wireless access networks. Three major approaches can be distinguished. In 1998, companies such as Ericsson, IBM, and Intel, formed a Special Interest Group in order to develop a wireless local area network technology that is cheap, while it offers different types of services, such as voice or best-effort traffic. Its code name was Bluetooth.

The second wireless technology for local area networks was Wireless LAN, initiated by the Institute of Electrical and Electronics Engineers (IEEE). It is commonly known as Wireless LAN. The basic standard IEEE 802.11 was defined in 1999.

The third initiative to create a substitution for the wired Ethernet, was the Broadband Radio Access Networks (BRAN) project. The planning of the standard started in 1991 and the HIgh PErformance Radio LAN version 1 (HIPERLAN/1) standard was approved in 1996. A further extension was accomplished in February 2000 with HIPERLAN/2. Compared to the other two proposals, Bluetooth and Wireless LAN, the HIPERLAN/2 technology supports high data rates and service differentiation for different types of Quality of Service.

However, the first products to get to market were based on Bluetooth and Wireless LAN technology. They offered little in terms of data rate and QoS support, but HIPERLAN/2 devices were simply not available. Today, there are a small number of HIPERLAN/2 devices available, but Wireless LAN is ubiquitous, while HIPERLAN/2 only exists in special environments. Therefore, the HIPERLAN/2 standard probably will not play an important role in the future.

Compared to the Bluetooth technology, Wireless LAN with its large number of extensions, as described in the next section, offers higher data rates and better QoS support. Bluetooth usually offers 1 Mbps and is restricted to an area of about 10 meters. Therefore, it is mainly used as a so-called Personal Area Network (PAN), where it connects the wireless keyboard and mouse to the computer, takes care of the communication between the laptop and the printer, or lets the user synchronize the address book of his PC and PDA.

Wireless LAN, on the other hand, is frequently seen as a nice, cost-effective and handy way to extend the Local Area Network. The high data rate and the low price already stimulate the discussion about a potential competition between 3G networks, such as UMTS, and Wireless LAN. Other opinions are, that a combination of the different technologies offers a by far greater opportunity. These networks of the 4th generation (4G), will allow very high data rates, low prices, anywhere-anytime connections, and support for QoS (see [AHP$^+$01], [rGPP02b], [rGPP02a], [rGPP03], [rGPP02c], [Inf02], [Inf01a], [Inf01b], [Inf03]).

The introduction of the Worldwide Interoperability for Microwave Access (WiMAX), a wireless Wideband Metropolitan Area Network based on the IEEE 802.16a standard [IEE03b], is supposed to cover large areas with data rates of 100 Mbps up to 1 Gbps, which even intensifies the idea of 4G networks. WiMAX will be used for two different purposes. First, it can be used to connect multiple Wireless LAN hot spots to a single high-speed wired Internet connection. This by far increases the flexibility in setting up Wireless LAN hot spots, since there is no need for individual Internet connections for each single hot spot. The second purpose for WiMAX is to directly connect the clients to the WiMAX network, wherever coverage available. Combined WiMAX and Wireless LAN (WiFi) client devices will soon be seen on the market.

In addition, handover mechanisms are introduced that allow vertical mobility, meaning seamless handovers between the different wireless technologies, such as UMTS and Wireless LAN ([SK98], [ZJD03]). These will allow a user to always connect to the wireless technology that provides the currently best service, in terms of bandwidth, QoS, or simply cost.

However, the currently available Wireless LAN devices do not support QoS in a way necessary for 4G networks. The next section explains the Wireless LAN standard family and the QoS extensions that were published. Later chapters then focus on the capabilities of these extensions in 4G environments. Our studies show that Wireless LAN can be configured in a way to support QoS in such scenarios once the QoS extensions are implemented.

2.3 Wireless LAN 802.11 Standard Family

This section summarizes the IEEE 802.11 standard family. While the initial standard defined the basic access mechanism, it soon turned out that this will not be sufficient for future deployments. A number of different task groups have been formed to enhance the protocols. Each of these extensions is shortly described in the following.

IEEE 802.11

The basic standard IEEE 802.11 *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications* was published in 1999. It defines the basic Medium Access Control (MAC) mechanism, which mainly consists of the Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol, and different Physical layers (PHY), such as Direct-Sequence Spread Spectrum (DSSS) and Frequency Hopping Spread Spectrum (FHSS). The frequency band is restricted to 2.4 GHz. The data rates 1 Mbps and 2 Mbps are supported, with only two modulation types, the Differential Binary Phase Shift Keying (DBPSK) and the Differential Quadrature Phase Shift Keying (DQPSK). In addition an Infrared (IR) Physical layer was specified.

IEEE 802.11a

The supplement *High-speed Physical Layer in the 5 GHz Band* was published in 1999 as well. It defines the Physical layer to be used in the 5 GHz frequency band. The basic modulation is Orthogonal Frequency Division Multiplex (OFDM), which splits the frequency band into smaller subcarriers that are simultaneously used to transmit data packets using lower data rates. Therefore, using a combination of subcarriers for the transmission of data, higher data rates can be achieved, while the subcarriers are less susceptible to inter-symbol interference. The modulation types Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), 16-Quadrature Amplitude Modulation (16-QAM), and 64-Quadrature Amplitude Modulation (64-QAM) are used on the subchannels and the Coding Rates 1/2 and 3/4 are implemented. Therefore, the data rates 6, 9, 12, 18, 24, 36, 48, and 54 Mbps become possible. Today, OFDM is the most promising candidate for future wireless communication systems.

IEEE 802.11b

The supplement *Higher-Speed Physical Layer Extension in the 2.4 GHz Band* defines two additional modulation schemes for the 2.4 GHz frequency band. Complementary Code Keying (CCK) and Packet Binary Convolutional Coding (PBCC) allow to increase the data rate to either 5.5 Mbps or 11 Mbps. This extension of the basic PHY layer is downward compatible, such that the devices using 1 Mbps and the new 11b devices with up to 11 Mbps can be operated simultaneously within a single Wireless LAN cell. The CCK modulation is mandatory while the PBCC modulation is only optional.

IEEE 802.11c

The bridging functionality necessary to implement the data exchange between the wireless and the wired medium is specified in the *Media access control (MAC) bridges* supplement of the standard. This extension focuses on improving the MAC layer for better bridging.

IEEE 802.11d

The task group d dealt with the problem of *Specification for operation in additional regulatory domains*. The main goal is to enhance the MAC protocol, such that it can be configured to better conform to the different regional regulations that can be found world-wide. It was published in 2001.

IEEE 802.11e

A major enhancement of the MAC protocol is specified in the *Medium Access Control (MAC) Quality of Service (QoS) Enhancements* amendment. The task group e defined extensions that allow the MAC protocol to distinguish between traffic types and to apply them to different priority levels. The parameters for the priority classes are not specified as fixed values, but can be adapted to the need of the Wireless LAN operator.

The IEEE 802.11e extension is the main focus of this work. Its capabilities to support different QoS levels in large-scale environments is evaluated in Chapter 5. The purpose of the QoS enhancement is to prepare the MAC protocol for the requirements of future 4G mobile networks.

IEEE 802.11f

In 2003 the *IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation* amendment to the Wireless LAN standard was published. Its main purpose is to define the Inter Access Point Protocol (IAPP), which specifies the communication between the access points in larger Wireless LAN environments. One of the tasks of IAPP is to support seamless handovers on the Data Link Layer.

IEEE 802.11g

Task group g published the *Further Higher Data Rate Extension in the 2.4 GHz Band* in 2003. It specifies the utilization of the Orthogonal Frequency Division Multiplexing modulation within the 2.4 GHz frequency band. The maximum physical data rate can be raised to a maximum of 54 Mbps as for the 5 GHz frequency band defined in IEEE 802.11a. Two different modes of operation were defined. One mode is downward compatible to the basic IEEE 802.11 standard, since header information is transmitted using 1 Mbps, while only the data portion is sent with a higher data rate. The second mode of operation is restricted to a pure 802.11g Wireless LAN network. In this mode, the header information is transmitted using the higher data rates, which by far reduces the overhead of the protocol.

IEEE 802.11h

Spectrum and Transmit Power Management Extensions in the 5 GHz band in Europe are defined in the IEEE 802.11h extension to the Wireless LAN standard. Its main purpose is to extend the PHY layer as defined in IEEE 802.11a in order to comply with the regulations for the 5 GHz frequency band in some European countries.

IEEE 802.11i

Security problems have been a matter of great concern since the first introduction of Wireless LAN. The basic security mechanisms have been found to be very weak. Therefore, the task group i was formed to define security mechanisms of higher sophistication. The 802.11i supplement *MAC Enhancements for Security* was released in 2004 and consists of a number of extensions to the basic Wired Equivalent Privacy (WEP) mechanism. As was expected, the mechanisms of the IEEE 802.1x standard were integrated into the new standard. It allows the RADIUS based access control and dynamic per session and per user encryption keys. However, since such an infrastructure is not an option for home environments, simpler and weaker mechanisms were added as well.

IEEE 802.11j

The definition of the *4.9 GHz - 5 GHz Operation in Japan* amendment to the standard was assigned to taks group j. As the 802.11h extension aims at the usage of the 5 GHz band in European countries, the 802.11j extension deals with the specific situation that the wireless technology is confronted with in Japan. The goal is a harmonization of the local regulations and the Wireless LAN 5 GHz PHY layer.

IEEE 802.11k

The supplement *Radio Resource Measurement of Wireless LANs* deals with the problem of how to retrieve the data necessary for the management and maintenance of a wireless network. The main focus lies on the type of data as well as the way in which it can be exposed to the outside environment.

IEEE 802.11REVma

The purpose of task group m is to define a *Revision 200x* of the 802.11 standard. It will be used to incorporate all the standard extensions that are approved up to a specific date. The goal is to create a single document that summarizes the amendments that are finished so far.

IEEE 802.11n

Enhancements for Higher Throughput are the topic of task group n. The goal is to define extensions to both 802.11 PHY layers and the 802.11 Medium Access Control layer to allow data rates of at least 100 Mbps. The PHY layer will be based on MIMO/OFDM (Multiple-Input-Multiple-Output / Orthogonal Frequency Division Multiplex) technology. It is assumed that data rates of up to 540 Mbps will be achievable.

IEEE 802.11p

Task Group p deals with the topic of *Wireless Access in Vehicular Environments*, which aims at inter-vehicle as well as roadside-vehicle communication. The amendment is supposed to at least support speeds of up to 200 km/h and communication ranges up to 1000 meters in the 5 GHz frequency band.

IEEE 802.11r

The amendment *Fast BSS-Transition* enhances the MAC layer in terms of handover delay within a single Extended Service Set (ESS), i.e. a layer two handover. IP handovers are not considered. However, security is to be kept at a high level. Decreasing the security in order to accelerate the handover is not wanted.

IEEE 802.11s

The definition of a Wireless Distribution System (WDS) and Extended Service Set Meshes is the topic of the *IEEE 802.11 ESS Mesh Networking* supplement. Its goal is to define self-configuring multi-hop topologies in order to improve the ad-hoc capabilities of the IEEE 802.11 wireless network.

IEEE 802.11.2

Amendment *P802.11.2 - Recommended Practice for the Evaluation of 802.11 Wireless Performance* deals with the definition of performance metrics, measurement methodologies, and test conditions. The goal is to create tools that allow measuring and predicting the performance of a Wireless LAN.

IEEE 802.11u

The *IEEE 802.11 Wireless Interworking with External Networks* amendment will extend the PHY and MAC layer in order to enhance the interworking of Wireless LAN with other networks.

IEEE 802.11v

Task group v deals with the standard supplement called *IEEE 802.11 Wireless Network Management*. The goal is to extend the capabilities of the PHY and MAC layer in order to support a better radio measurement, which in turn allows to better interface to the upper layers for managing 802.11 devices in wireless networks.

Summary

This section listed all extensions that have been or will be defined for the basic Wireless LAN standard. As of the time of writing, just a small number of these amendments have been finished: IEEE 802.11a, IEEE 802.11b, IEEE 802.11c, IEEE 802.11d, IEEE 802.11f, IEEE 802.11g, IEEE 802.11h, IEEE 802.11i, and IEEE 802.11j. Devices implementing these extension are available on the market. All the other working groups are still active.

The most important extension for our studies is, however, the IEEE 802.11e amendment. It supplements the original MAC protocol with support for Quality of Service. The working group expects to finish its work in the second quarter of 2005.

3 Wireless Challenges

> Multimedia? As far as I'm concerned, it's reading with the radio on! Rory Bremner (1961)

We want to study the capabilities of Wireless LAN in the context of 4G networks. The main goal is to understand what kind of traffic or traffic mix can be supported in large-scale environments. However, there are different traffic types that have to be considered separately. In this chapter the different traffic types, that have to be taken into account, are discussed.

Section 3.1 introduces the main traffic types that can be distinguished in today's networks. These traffic types exhibit different Quality of Service demands on the network. While an FTP download might be best if the average data rate received is optimized, this is not an important factor for voice traffic. Voice applications produce rather low data rates while the end-to-end delay has to be kept within a certain level in order to keep the voice transmission at an acceptable quality.

Section 3.2 introduces the different QoS demands of the traffic types in general. However, simple measures, such as end-to-end delay or packet loss, are not sufficient for the evaluation of voice or video quality. Section 3.2, therefore, introduces more advanced mechanisms to assess the user experienced quality of real-time traffic as well.

Finally, Section 3.3 summarizes the traffic types and the quality assessment methods that are considered in later chapters.

3.1 Traffic Types

Basically, IP data traffic can be distinguished in terms of reliability. If the Transmission Control Protocol (TCP) is used, the connection is reliable. TCP takes care of the retransmission of lost data packets and it assures that the packets are received in the correct sequence. This type of transmission is used in applications like the World Wide Web (WWW) or FTP, which rely on the correct reception of all data packets. In addition, TCP implements a flow control mechanism which tries to utilize the whole available bandwidth. It increases the rate of the data transmission as much as possible, such that, in the case of lost packets, the data transmission rate is decreased and overload situations can be avoided.

On the other hand, the User Datagram Protocol (UDP) is unreliable. It does not perform retransmission of lost packets and it does not account for packets received in the wrong sequence. However, if real-time applications are performed, it is mandatory to receive the data packets within a certain time limit, while a small amount of lost packets is quite acceptable. Retransmissions would only lead to an increased delay of the data transmission. Therefore, the UDP protocol is the better alternative for real-time applications, e.g. voice or video transmissions.

In the remainder of this section, the four different traffic types WWW, FTP, voice, and video are described in further detail.

3.1.1 File Transfer Protocol (FTP)

The File Transfer Protocol (FTP), defined as an IETF standard in [PR85], specifies a way to transfer complete files between two computers. A set of data formats and commands is standardized in order to allow a user to copy a file from one computer to another. It is implemented on top of TCP [Pos81]. Therefore, during file transmission, the FTP protocol tries to utilize the whole available data rate.

In the remainder of this work, FTP file transfer is frequently used as a worst-case scenario for a Web user. As discussed in the next section, a Web user frequently transfers files from the Web server to the local computer, while these downloads are intermingled by idle periods. An FTP user that continuously starts a download after the last file transfer completes, represents the case where the idle periods of a Web user activity is ignored, but constant activity is assumed. This allows us to perform comparably short simulation runs, but to still receive significant results that can be used to assess the quality of any kind of non real-time data transmission.

3.1.2 Hypertext Transfer Protocol (HTTP)

The Hypertext Transfer Protocol (HTTP) is the underlying protocol of the World Wide Web (WWW) application. Version 1.1 is defined in RFC 2068. It defines the way in which the data is transferred between a Web server and a Web user. Figure 3.1 shows an overview of the protocol layers involved in Web users' activities. On the top, different client activity phases are shown. Within each activity phase, a number of Web sessions is performed. A Web session is a number of Web pages that a user downloads within its activity phase ([TGSL01]).

The requested Web pages can be represented by an On/Off process, where each On phase corresponds to the download of a single Web page, while the Off phases define the time between the download of consecutive pages, meaning for example the reading time. A single Web page consists of a number of objects, the main and the inline objects, which are transmitted consecutively within a maximum number of four simultaneous TCP connections. Each such TCP connection then consists of a number of TCP packets depending on the size of the transmitted objects. Lower layers such as Wireless LAN might further fragment the TCP packets.

A complete description of this Web source traffic model can be found

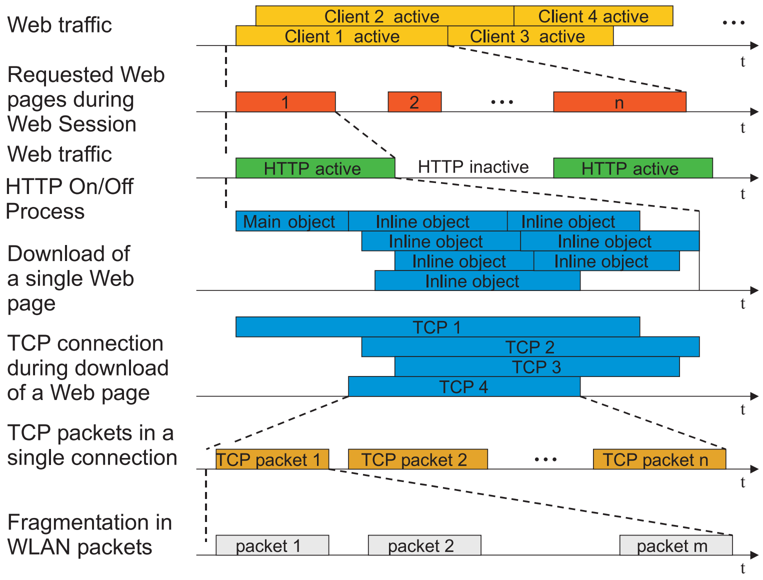

Figure 3.1: *Web source traffic model*

in [TGSL01]. This includes a summary of the statistical properties of the Web traffic, such as the distribution of the session duration or the number and size of the inline objects.

However, since the statistical properties of Web sessions are highly variable, very long simulation runs have to be performed in order to retrieve statistically relevant results. Most of the time, it is not a practical solution to simulate the exact behavior of even just a small number of Web users. The bandwidth requirements of a single Web user is rather small, i.e. in the order of tenths of Kbps. Nevertheless, if the whole bandwidth of a Wireless LAN system is to be consumed by the simulated Web users, a huge number of simultaneously active Web users have to be taken into account, and the simulation time exceeds practically feasible amounts. Therefore, in the remainder of this work, FTP users are simulated to retrieve performance results for best-effort traffic as was already explained in the last section.

3.1.3 Voice data transmission

The most widespread way to transmit speech in legacy circuit switched networks is to take an eight bit speech samples every 125 microseconds, leading to an overall bandwidth requirement of 64 Kbps. In circuit switched networks, however, a channel between the two transmitting stations is set up at the beginning of a call and bandwidth is exclusively reserved for its whole duration. Therefore, the 64 Kbps channel in both directions is reserved all the time and can not be used by other applications.

Packet switched networks, however, do not reserve bandwidth for for the whole duration of a single connection. They allow other applications to share the same medium. Therefore, it is desired to reduce the bandwidth requirement of a voice application as much as possible. And indeed, voice data allows a great amount of optimization in terms of bandwidth requirement. The most obvious one is the silence detection and suppression. During a voice call, usually only one party speaks at at time. Therefore, half of the time there is silence and no data needs to be transmitted in one direction.

On the other hand, modern coding theory allows to even further decrease the required bandwidth of a simple voice call. A number of different codecs, pairs of coder and decoder, were introduced. They allow to remove redundancy that is immanent in human speech. Table 3.1 summarizes a number of them. They mainly differ in terms of required data

Codec	Data Rate [Kbps]	Frame Size [ms]	Data Size [Bytes]
G.711 PCM [IT93a]	64	0.125	1
G.726 ADPCM [IT96a]	32	0.125	0.5
G.729 CS-ACELP [IT96b]	8	10	10
G.723.1 MP-MLQ [IT96c]	6.4	30	24
G.723.1 ACELP [IT96c]	5.3	30	20

Table 3.1: *Compression algorithms and bandwidth requirement*

rate, frame size, and data size. The frame size specifies the interarrival time between consecutive packets, i.e. for G.711 [IT93a] every 125 microseconds a single voice packet is sent, while for the G.723.1 ACELP codec [IT96c] the interarrival time between two packets is 30 milliseconds. The frame size of most codecs can, however, be varied by the application. The G.711 codec, for example, is often configured with a frame size of 4 ms in order to reduce protocol overhead. Later chapters show, that the frame size has the most important effect on the performance of voice transmissions in Wireless LAN networks.

As described in the introduction of this section, the User Datagram Protocol (UDP) as defined in [Pos80] is used to transport the data packets from the source to the destination. The UDP protocol is unreliable in terms of packet loss, i.e. lost packets are not recovered by the protocol. Higher layer protocols have to get along with the lost information or they have to perform the retransmission procedure by themselves.

In case of voice data, the higher layer protocol is *A Transport Protocol for Real-Time Applications (RTP)*, also known as the *Real-time Transport Protocol* [SCFJ03]. It takes care of the resequencing of the data packets in cases where they are not received in the correct order. This might be caused by a link failure and a resulting change of the routing in the backbone network. RTP does not implement retransmissions in case of lost packets. The reason is that the most important factor for an acceptable quality of a voice transmission is delay. Retransmissions, however, lead to an increase of the delay. Packet loss, at least of a small amount, is not an issue. Later in this chapter, the different quality assessment procedures are covered in more detail.

3.1.4 Video data transmission

Video telephony, online video conferencing, and video streaming are the most promising applications for future wired and wireless networks. However, the unencoded transmission of videos is not an option due to the immense bandwidth requirements. As for voice data, a number of different video codecs have been developed in to lower the bandwidth requirements. The International Telecommunication Union (ITU) and its ITU Telecommunication Standardization Section (ITU-T) defined video source coding algorithms that help to reduce the required data rate due to the reduction of video immanent redundancy ([IT93b], [IT95], [IT02]).

The most important standard for our studies is the H.263, which defines *Video Coding for Low Bitrate Communication*. It specifies a hybrid of inter-picture prediction to utilize temporal redundancy and Discrete Cosine Transform (DCT) of the residual prediction error to reduce spatial redundancy. After the DCT coding, the prediction error is quantized and the resulting symbols are variable-length-encoded and transmitted.

Therefore, H.263 [IT95] encoded video stream consists of three types of frames. The *I-Frames* (intra) are solely intra-coded frames and represent a whole picture. The *P-Frames* (predicted) follow the I-Frames or P-Frames and contain only the data that has changed from the preceding I-Frame or P-Frame. Finally, the *B-Frames* (bi-directional) rely on the frames preceding and following them. They contain only the data that has changed from the preceding I- or P-Frame or from the data in the next I- or P-Frame.

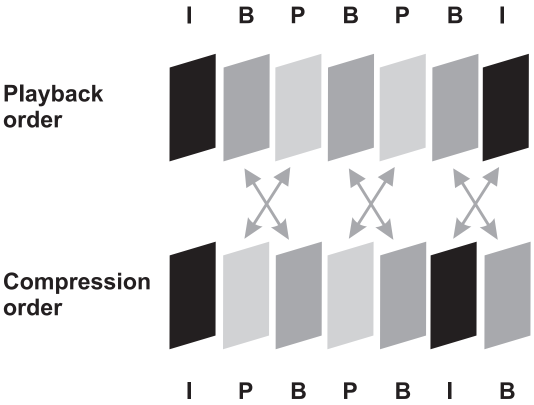

Figure 3.2: *Playback and compression order of a GoP*

The different frame types are transmitted in recurring sequences of frames, also known as Group of Pictures (GoP). Figure 3.2 shows an example of the transmission of an H.263 encoded video stream. As B-Frames need the preceding and the following frame to build a whole picture, their position in the compressed video stream differs from their playback position. B-Frames are usually not used in real-time video streams such as video conferences. This is due to the fact that the decompressor would require access to two frames in advance. Hence, a buffer is needed and its temporal length would add to the transmission delay. Moreover, B-frames are harder to calculate than P-Frames leading to a higher CPU workload in good cases and to a longer time to compress the video stream in bad cases. For additional information on video compression see [NH95], [Ric99].

Table 3.2 summarizes the parameters that are used for the videos simulated in this work. Both standard resolutions Common Interframe

Codec	Format	Resolution [Pixels]	Frame Rate [Frames per second]
H.263	CIF	352 x 288	25
H.263	QCIF	176 x 144	25

Table 3.2: *H.263 codec parameters*

Format (CIF) and Quarter CIF (QCIF) are considered. The frame rate can be set in the range of one to a maximum of 30 frames per second. In our case, 25 frames per second are assumed, which allows fluent videos. The different video formats are chosen depending on the platform that is used. For TV like screens the CIF resolution is chosen, while smaller screens, as for example in Personal Digital Assistants (PDA) or Cellular Phones, utilize the QCIF format. Online video conferences put the highest demand on the network due to the real-time nature of the data stream. Therefore, no B-Frames are used in our studies.

As for the voice traffic, UDP is used on top of the IP layer. Again, no retransmissions are performed and the RTP layer has to take care of the resequencing of the received packets. Lost packets decrease the quality of the video stream. If the delay of a packet is too large, such that the time of its scheduled playback time has passed at the arrival of the packet, the received video data is dropped. Later in this section, the QoS parameters for measuring the quality of the video transmission are discussed.

3.2 Quality of Service Assessments

The last section introduced the various traffic types that have to be analysed differently, when evaluating the QoS capabilities of a Wireless LAN network. It was pointed out that four types of traffic have to be analyzed individually. This section describes the measures and procedures to assess the user experienced Quality of Service.

3.2.1 File Transfer

Users that perform a file transfer using the File Transfer Protocol (FTP) are solely interested in the download time. This is directly influenced by the average data rate that the user experiences, which in our case is defined as the ratio between the amount of transmitted data and the time from connection setup until completion of the download. Packet loss and transmission delays lower the achieved data rate. The underlying TCP protocol and its rate control mechanism adapt the sending rate appropriately, such that an explicit assessment of the packet loss and delay is not necessary in this case.

It is worth mentioning that the size of the transferred file has an impact on the achievable data rate, since the TCP protocol and its rate control mechanism start with a low transmission rate and then probes the available bandwidth by increasing the transmission rate according to the Slow Start mechanism. In addition, the delay due to the connection setup phase lowers the achievable data rate as well. Therefore, very small files can never experience a high data rate, since the rate control mechanism is not able to utilize the whole available bandwidth in such a case. Therefore, files with varying sizes of one Kilobytes up to a maximum of ten Megabytes are considered in later chapters. The average data rate received by the FTP user is the most important factor and is used to derive the results.

3.2.2 Web Traffic

The Quality of Service received by a user surfing the Web is much harder to measure. The average bandwidth requirement of a single Web user is around 10 Kbps based on the model as described in Section 3.1.2. Therefore, a single Web user can never utilize the full bandwidth that is provided by the network. Hence, the average data rate received by a single user is not of great value when Web traffic performance is to be analyzed.

The main interest of a Web user is the average page download time, that is the average time required for the download of a single Web page. The higher the load on the network, the longer the delays and the higher the packet loss become. This directly influences the delay of the downloaded Web page. The underlying TCP protocol assures the correct reception of all the data packets. However, the retransmissions of lost packets add to the overall delay. Therefore, the measure of choice for the assessment of the Web traffic performance is the average page download time experienced by a user.

As explained above, the Web source traffic model leads to high variations in terms of page size and other parameters, which makes very long simulation necessary. When this kind of simulation becomes unfeasible, we simulate FTP users in order to evaluate the QoS received by best-effort users in the remainder of this work.

3.2.3 Voice Traffic

The classical way to assess the quality of a transmitted voice stream is to measure different statistics, e.g. delay, jitter, or packet loss. These parameters are directly taken from the TCP or IP packet stream. No real voice stream has to be transmitted, but it is only necessary to simulate a characteristic voice stream, i.e. a packet of a certain size has to be sent repeatedly after constant amounts of time.

The Study Group 12 of the ITU-T [Cov01], therefore, specified key performance parameters and target values that adapt to the different voice applications. In this work, three different voice applications are

Application	Key performance parameters and target values		
	One-way delay	Delay variation	Information loss (**)
Conversational voice	< 150 msec preferred (*) < 400 msec (*)	< 1 msec	< 3 % packet loss ratio (PLR)
Voice messaging	< 1 sec for playback < 2 sec for record	< 1 msec	< 3 % PLR
High quality streaming audio	< 10 sec	< 1 msec	< 1 % PLR

(*) Assumes adequate echo control

(**) Exact values depend on specific codec

Table 3.3: *ITU-T audio parameter recommendations*

distinguished. Conversational voice is the only symmetric application, where data flows in both directions simultaneously. Voice messaging as well as high quality streaming audio is primarily one-way. Conversational voice data ranges in data rate from about 4 Kbps to 64 Kbps. Voice messaging has similar requirements with about 4 Kbps to 32 Kbps. High quality streaming audio is clearly the most demanding application in terms of data rate ranging from 16 to 128 Kbps.

In terms of QoS, the most demanding voice application is definitely conversational voice. The one-way delay limits of 150 msec preferred and 400 msec maximum are considerably short. Even a packet loss ratio of less than three percent is hard to achieve in wireless environments.

However, using such objective measures has one major drawback. As the second footnote of Table 3.3 already indicates, some of the recommended parameters depend on the speech codec that is used. Some codecs do not exhibit a perceivable quality degradation even in the case when the packet loss reaches the three percent packet loss ratio. On the other hand, three percent packet loss can lead to incomprehensible speech. Hence, the simple evaluation of delay, jitter (delay variation), or packet loss is not a good choice.

Therefore, a different approach is chosen in this work. The goal is to measure the quality of the voice stream at the receiver side objectively. In order to do so, the ITU-T developed the PESQ algorithm in its P.862 recommendation *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs* [IT01].

The overall procedure of PESQ is shown in Figure 3.3. It measures the end-to-end speech quality of a one-way transmission. In order to do so, a reference voice stream is passed through the system and the degraded signal as experienced at the receiver side is created. The PESQ algorithm takes the reference and the degraded signal as input and measures the level of degradation.

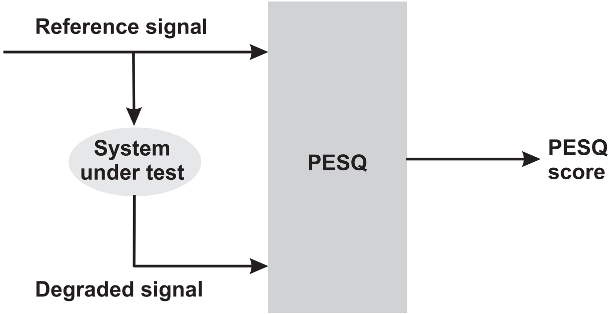

Figure 3.3: *PESQ procedure overview*

The resulting PESQ score is then used to derive a Mean Opinion Score (MOS) as defined in [IT98], [IT96d], and [IT96e]. As shown in Figure

3.4 the MOS values range from 1.0, meaning *not recommended*, to a maximum of 4.5, which represents *very satisfied* users. The method of using the PESQ algorithm and the MOS value mapping allows to derive an objective measure for the transmitted voice quality independent of the considered voice codec.

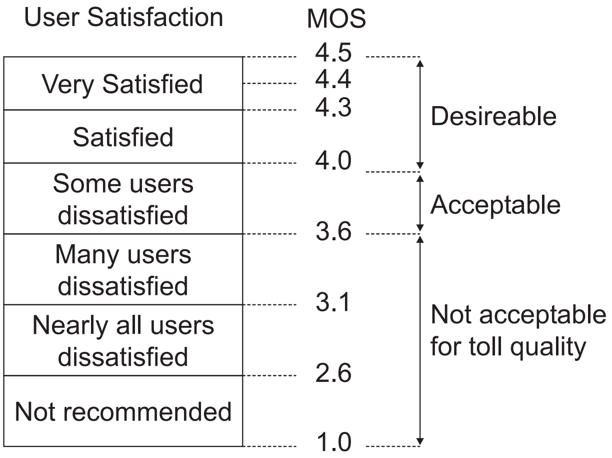

Figure 3.4: *Mean Opinion Score (MOS)*

The PESQ algorithm was designed in a way that it matches the subjective opinion of human test persons as much as possible. It takes into account features such as filtering, variable delay, coding distortions, and channel errors. Therefore, it is a lot more complicated to implement and to measure than parameters such as delay or jitter. However, PESQ and MOS are the better mechanisms to derive the user experience quality of a voice stream.

It should be mentioned that some proposals for an improvement of the transmission of voice streams exist in the literature. In [HRW03] the authors, for example, show that the packets of a single voice stream can be distinguished in terms of their importance for the quality of the received signal. The goal here is to assure that packets with a higher importance receive a better quality than those of low importance. However, since such methods are highly complicated and can not be seen in any practical environments so far, thea are not accounted for in our studies.

3.2.4 Video Traffic

In the last section, the different approaches to measure voice quality at the receiver side were discussed. It was shown that there are simple measures, such as delay, jitter, and packet loss, that can easily be retrieved from simulation runs. However, these simple approaches are not always adequate to measure the quality of the voice stream as experienced at the receiver's side. More advanced mechanisms can be deployed. It is much more complex to retrieve the results, but the measured QoS directly states the level of user satisfaction.

The same holds for video traffic. The Study Group 12 of the ITU-T [Cov01] recommended some hard limits for the parameters delay and packet loss. This is the classical approach. Table 3.4 summarizes the key performance parameters for video transmissions. While the videophone application is two-way, there also exist recommendations for the one-way video streaming applications. Clearly, the interactive videophone has the higher demand. Data rates for both applications range from 16 Kbps to about 384 Kbps, but the videophone has the same short delay targets as the interactive voice call. The goal of our studies is to support the videophone application.

Again, as for voice traffic, the video codec plays an important role on the quality of the received video stream as well. The simple parameters, delay and packet loss, do not allow to assess the quality as experienced by the user. Therefore, more advanced techniques have been proposed

Application	Key performance parameters and target values	
	One-way delay	Information loss
Video phone	< 150 msec preferred < 400 msec limit	< 1 %
One-way	< 10 sec	< 1 % PLR

Table 3.4: *ITU-T video parameter recommendations*

for video traffic by the ITU-T. The proposed measure is the Peak Signal to Noise Ratio (PSNR), a subjective interpretation of the quality of a transmitted video stream. PSNR is a derivative of the well known Signal-to-Noise Ratio (SNR) and is the most widespread technique for the quality assessment of video.

The PSNR is calculated image by image. It compares the maximum possible signal energy to the noise energy, which results in a higher correlation with the subjective quality perception than the conventional SNR. The definition of the PSNR of source image s and destination d is given by the following equations (see [NH95], [RJ91]). More information about the PSNR calculation can be found in [Ric99], [FSR04].

$$\begin{aligned} PSNR(s,d) &= 10\log_{10}\frac{V_{peak}^2}{\sqrt{MSE(s,d)}}[dB] \\ &= 20\log_{10}\frac{V_{peak}}{\sqrt{MSE(s,d)}}[dB] \\ V_{peak} &= 2^k - 1, \text{k bit color depth} \\ MSE(s,d) &= \text{mean square error of } s \text{ and } d \end{aligned}$$

The PSNR values do not directly correspond to MOS values [Cis04]. However, the heuristic mapping as shown in Table 3.5 is used in our

Quality	PSNR [dB]	MOS
Excellent	≥ 37	5
Good	[31;37)	4
Fair	[25; 31)	3
Poor	[20; 25)	2
Bad	< 20	1

Table 3.5: *Heuristic PSNR to MOS mapping*

studies [IT96d]. This allows to give a computational approximation of the subjective human impression of the video stream, similar to the voice stream evaluation.

3.3 Traffic Categories in the Remainder

The last section described the different traffic categories that have to be studied when evaluating the capabilities of a Wireless LAN network in the context of a 4G system. In this section, a short summary of the traffic types as they are used for the simulation studies in later chapters is given.

The main traffic type for the evaluation of best-effort performance is FTP. An FTP user is simulated in a way that it consecutively downloads files of varying sizes from an FTP server. The measure of choice is the average bandwidth received. Even though Web traffic is experienced much more often in today's networks, it is less appropriate to measure the received QoS for best-effort users as explained above. Therefore, Web traffic is only considered in some of the simulation scenarios.

IP telephony is simulated using a 16 bit, 8 kHz mono wave audio sample with the length of 90 seconds. This corresponds to the mean duration of a telephone call in mobile communication networks. The wave audio sample is encoded using the G.723.1 codec with a coding rate of 5.3 Kbps and 6.3 Kbps and VAD (Voice Activity Detection) disabled. The 90 second G.723.1 audio stream consists of exactly 3000 voice frames. The size of a single voice frame on the IP layer is 20 Bytes for the G.723.1 (5.3 Kbps) and 24 Bytes for G.723.1 (6.3 Kbps). The frame size is set to 30 ms. G.723.1 is a low quality voice codec, but it shows an acceptable performance in Wireless LAN reference scenarios as shown in later chapters.

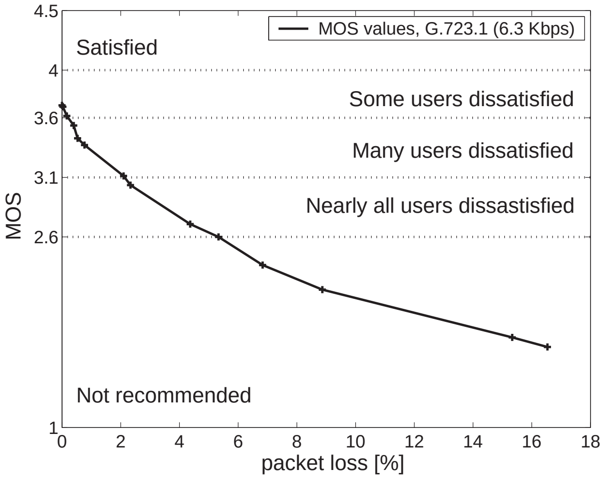

Figure 3.5: *MOS for G.723.1 with 6.3 Kbps*

Figure 3.5 shows the effect of packet loss on the MOS rating of a G.723.1 coded audio stream when 6.3 Kbps are used. It is easy to see that even a small percentage of packet loss leads to a dramatic degradation of the received voice quality. The main reason is that the G.723.1 codec has a large coding factor, i.e. a high level of compression. The codec does not

Compression Algorithm	Bit rate [Kbps]	MOS Score
G.711 PCM	64	4.1
G.729 CS-ACELP	8	3.92
G.723.1 MP-MPQ	6.3	3.9
G.723.1 ACELP	5.3	3.65

Table 3.6: *Maximum achievable MOS values for different voice codecs*

only remove the redundancy, but it leads to a degradation of the speech quality. This can be seen for the case where no packet loss occurs. Here, the maximum MOS reaches about 3.7, which is quite acceptable, but already leads to some dissatisfied users. Depending on the compression algorithm, the maximum achievable MOS differs as shown in Table 3.6.

However, as we see later, using the codec with the highest achievable MOS value does not lead to a better performance of the system in terms of QoS support. Using the G.723.1 codec [Kab03] is shown to be the best solution in case of Wireless LAN networks in Chapter 5.

In the case of voice traffic, the traffic pattern is rather simple to simulate. Depending on the codec, the packet and frame sizes differ, but are constant for the whole duration of the call. Therefore, the simulation just has to model a voice source as a packet generator, that sends a packet to the receiver every frame size milliseconds with a constant packet size. The situation changes once video applications are considered.

The data rate that is required for the transmission of a video stream, is not at all constant over the whole time period. The data rate of a video changes rather rapidly depending on the video stream itself. Therefore, for video simulations a real video has to be chosen and its properties have to be analyzed. In our case, eight different video sequences where taken. Four two-minute sequences are taken randomly from the movie *Red Rock West (1994)* and another four two-minute sequences from the movie *Kill Bill: Vol. 1 (2003)*. These test sequences were analyzed in terms of their statistical properties. The results for the best-case and worst-case sequences are summarized in Table 3.7. Best-case and worst-case refer to the statistical properties of the video. The less variable in terms of frame size the video is and the lower the average data rate it requires, the better it is suited to be transferred on a packet based network. On the other hand, the more the frame sizes of the video stream vary, the higher is the demand on the system.

Video	Format	Size [KBytes]	Number of Packets	Average Frame Size [B]	Standard Deviation [B]
A	CIF	3792.8	4179	1237.5	942
B	CIF	1165.7	3018	469.5	361
C	QCIF	1488.8	3062	361	263
D	QCIF	494.4	3001	138	103

Table 3.7: *Properties of the best-case and worst-case video sequences*

The video statistics were calculated using the frame size information of the different video streams. The sample mean is derived from the frame sizes x_k by

$$m = \frac{1}{N}\sum_{k=1}^{N} x_k.$$

Using the sample mean m, the unbiased sample variance s_{N-1}^2 and the standard deviation s_{N-1} can be calculated by

$$\begin{aligned} s_{N-1}^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - m)^2, and \\ s_{N-1} &= \sqrt{s_{N-1}^2}. \end{aligned}$$

These statistics allow to receive the nessessary information about the variability of a given video stream.

Therefore, Video A represents the worst-case video sequence in the case of CIF format. The averaged frame size with 1237.5 Bytes is roughly three times higher than for the best-case video sequence B. The difference between these two CIF formatted videos in terms of variability is just as high.

Similar results can be found by studying the statistical properties of the QCIF formatted video sequences. The worst-case is marked as video C. Again the required data rate is about three times higher than for the best-case video D. The variability of video C also reaches levels almost three times as high as for video D.

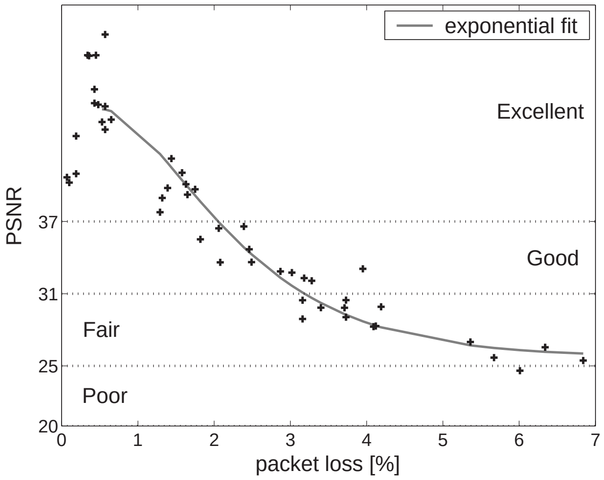

Figure 3.6: *PSNR values depending on the packet loss probability*

Figure 3.6 shows the effect of the packet loss on the Peak Signal to Noise Ratio (PSNR). This plot was produced using video A, the worst-case CIF formatted video sequence. The dotted lines in the figure indicate the different MOS levels. It can be seen that the video quality is excellent as long as the packet loss does not exceed a level of two percent. The quality that the user experiences is still good, even for value around three percent. Then, the quality reaches *fair* and drops only gradually with an increased packet loss.

It is important to mention that the quality of a received video can not be measured only depending on the packet loss rate. Consider a video where an I-Frame is lost. The I-Frame, as explained before, carries the information about the whole picture, and thus forms the basis for a whole GoP. Therefore, the whole GoP is disturbed and the effect of the lost I-Frame is clearly perceivable.

This fact can be seen in Figure 3.7. Here, two different screenshots are shown from video A. On the left-hand side, the original image is displayed. On the right-hand side, a disrupted image is shown. In this case, the I-Frame was lost, such that the codec filled the missing image with a simple background color. Then, the P-Frames arrive one at a time, but they do not contain information about the whole image, but they just describe the changes in the image. In our case, the area around the moving person in the picture has been transmitted, and thus can be displayed by the codec. The rest of the image can not be shown correctly.

This example shows that the packet loss as the only measure for the quality of the received video is not sufficient. The type of packet that has been lost, needs to be taken into account as well. This, however, is difficult to measure in terms of packet loss. The PSNR or MOS value on the other hand measures the received quality perceived by the user and, therefore, implicitly takes the importance of the lost frames into consideration.

Finally, Figure 3.8 shows two series of frames from video A. The leftmost frame always shows the image taken from the original movie sequence. Therefore, its MOS value is 5. The other two frames show different levels of distortion. The frame in the middle is taken from a

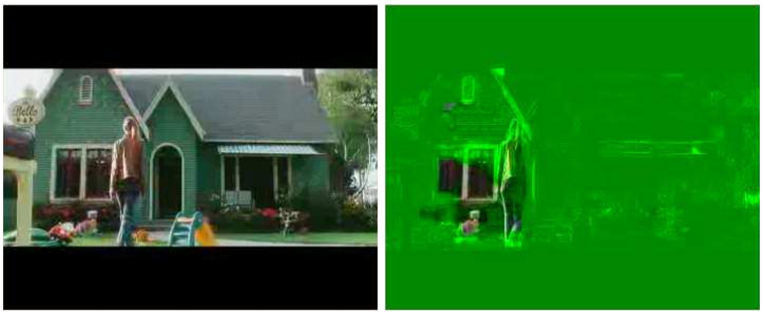

Figure 3.7: *Distortion caused by lost I-Frame*

degenerated video with 2.4 percent packet loss and a MOS value of 4. The distortions can be perceived already, but the overall movie still has an acceptable quality.

The leftmost frames in Figure 3.8 were taken from a sequence with a packet loss of 3.3 percent and a MOS value of 2. The quality in this case is *poor*. Large areas of lost video information are experienced and the quality of the received movie sequence is not acceptable for a human user.

These examples proof that the PSNR calculation and the heuristic MOS mapping provide realistic results of the user experienced quality of a transferred video stream.

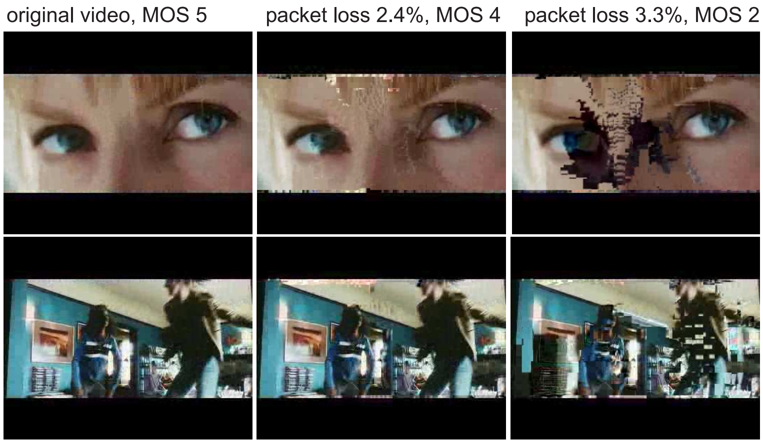

Figure 3.8: *PSNR values depending on the packet loss probability*

4 PHY - WLAN Physical Layer

> A technique succeeds in mathematical physics, not by a clever trick, or a happy accident, but because it expresses some aspect of a physical truth. Sir Oliver Graham Sutton

The Physical layer is the layer one of the ISO/OSI protocol stack. Its main purpose is to define the low-level specification of the bit-wise data transmission on a given medium. In the case of Wireless LAN these specifications include e.g. the frequency of the channel, the signal strenth, or the modulation. In the following, an overview of the Wireless LAN Physical layers as defined in the standard and its extensions is given.

Three different Physical layers (PHY) are defined in the IEEE 802.11 standard [IEE99a]:

- *Frequency-Hopping spread spectrum (FHSS) PHY specification for the 2.4 GHz Industrial, Scientific, and Medical (ISM) band*,
- *Direct sequence spread spectrum (DSSS) PHY specification for the 2.4 GHz band designated for ISM applications*, and
- *Infrared (IR) PHY specification.*

However, only the DSSS Physical layer found its way to the market, while the other two specifications are of minor practical relevance.

Originally, DSSS used the 2.4 GHz frequency band as the transmission medium and differential phase shift keying (DPSK) as the modulation technique. Two different flavors of DPSK are specified, Differential Binary Phase Shift Keying (DBPSK) for data transmission at 1 Mbps and Differential Quadrature Phase Shift Keying (DQPSK) at 2 Mbps. The protocol extension IEEE 802.11b [IEE99b] added two more modulation techniques to the original standard; Complimentary Code Keying (CCK) developed by Intersil and Packet Binary Convolution Coding (PBCC) first introduced by Texas Instruments. They were the first to allow higher data rates of up to 22 Mbps.

At the same time, the IEEE 802.11a standard [IEE99c] was defined. It uses the 5 GHz frequency band and Orthogonal Frequency Division Multiplex (OFDM) modulation. This allows for a further increase in the maximum data rate. IEEE 802.11a was the first Wireless LAN standard that could support data rates of up to 54 Mbps. Later the IEEE 802.11g standard specified the use of OFDM or CCK/OFDM for the 2.4 GHz band as well. Therefore, the maximum data rate in the 2.4 GHz band could also be improved to 54 Mbps.

In the following, the various modulation techniques are explained in more detail and the advantages and disadvantages of different approaches are discussed. However, modulation is not the only important factor when investigating the performance of a wireless system. There is a whole set of factors that influence the quality of the wireless link, and thus have an impact on the packet error rate and on the performance. Therefore, an overview of the most important factors is given in Section 4.2. Finally, Section 4.3 summarizes the settings that have been used for the performance analyses in later chapters.

4.1 Radio Transmission Techniques

The different PHY layers defined in the Wireless LAN standards serve the same tasks. They provide a frame exchange functionality to take care of the communication with the upper Medium Access Control (MAC) layer and indicate the MAC layer of the current medium status (busy/idle). Each PHY layer uses its own modulation techniques, channel settings, or frequency bands. However, all of them work with the same MAC layer, which is described in the next chapter.

In the following the different PHY layers are explained in greater detail. Geographical differences in the way the PHY layers have to work due to regulations are also covered.

4.1.1 FH-CDMA

The *IEEE 802.11 Frequency Hopping Spread Spectrum (FHSS) Physical Layer* can deliver 1 Mbps and 2 Mbps data rates. It uses frequency hopping spread spectrum in the 2.4 GHz frequency band. Depending on the geographical region, the frequency band is divided into a number of different channels. In North America and most of Europe the frequencies from 2.402 to 2.480 GHz are used, while the operation in Japan is restricted to the frequency band from 2.473 to 2.495 GHz. Each channel is 1 MHz wide, such that there are 79 separate channels in North America and most of Europe and only 27 channels in Japan.

The hopping sequence is determined by the access point, and the clients automatically synchronize to the correct hopping sequence. The number of hopping sequences also depends on the geographical region. In North America and most of Europe the standard specifies 78 different sequences and 12 sequences for Japan. They are chosen in order to avoid interference between co-located access points. Finally, the hop rate and the hop distance can be configured. The standard defines a minimum hop rate depending on the region, e.g. 2.5 hops per second in North America, and a minimum hop distance in the frequency domain of either 6 MHz for North America and most of Europe or 5 MHz for Japan.

Gaussian Frequency Shift Keying (GFSK)

Frequency Hopping Spread Spectrum delivers 1 Mbps and 2 Mbps data rates. In both cases Gaussian Frequency Shift Keying (GFSK) is utilized. GFSK is simply Frequency Shift Keying, but the input is first passed through a Gaussian Filter. Therefore, the binary data is not represented in the form of square pulse signals, but Gaussian pulses instead.

FSK thens represent the binary data in the form of different frequencies. For the 1 Mbps data rate two-level GFSK is used. Here, a binary 1 is delivered as a signal with higher frequency than the center operating frequency of the channel. A binary 0 instead is represented by a signal with a lower frequency than the center operating frequency. The standard defines a nominal frequency deviation of 160 KHz. The two-level GFSK is shown in Figure 4.1. The binary 1 is represented by a higher frequency than the binary 0.

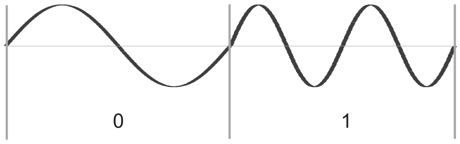

Figure 4.1: *Two-level Gaussian Frequency Shift Keying*

The 2 Mbps data rate is consequently implemented by using four-level GFSK, i.e. four different frequencies transmit two binary values simultaneously. This is shown in Figure 4.2.

The transmit power of the Wireless LAN devices in FHSS mode is restricted to a maximum output power of 100 milliwatts. This is the

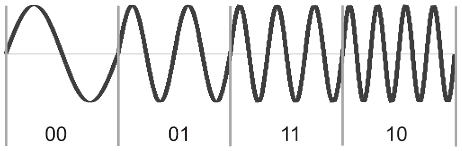

Figure 4.2: *Four-level Gaussian Frequency Shift Keying*

same for North America and Europe.

However, the Wireless LAN devices using FHSS mode do not have a great market share. This is mostly due to the higher data rates that DSSS devices offer.

4.1.2 DS-CDMA

The *Direct sequence spread spectrum (DSSS) PHY specification for the 2.4 GHz band designated for ISM applications* was the second initial 802.11 Physical layer implemented in Wireless LAN solutions. It uses a special type of Code Division Multiple Access (CDMA) to allow multiple users to share the single medium. However, there is a big difference between CDMA and DSSS. While CDMA uses multiple orthogonal spreading sequences in order to enable multiple users to operate at the same frequency simultaneously, DSSS only supports one single spreading sequence, which is used by all clients.

Therefore, DSSS does not allow simultaneous transmissions in the same frequency. The spreading is only used in order to increase the robustness of the transmission against other sources of interference, e.g. Bluetooth radio transmission. The Medium Access Control (MAC) protocol controls the access sharing of the clients. Simultaneous transmission is only allowed in different channels. The number of channels depends on the geographical region. In North America only 11 channels are de-

Channel number	Frequency (GHz)	North America	Europe	Japan
1	2.412	√	√	
2	2.417	√	√	
3	2.422	√	√	
4	2.427	√	√	
5	2.432	√	√	
6	2.437	√	√	
7	2.442	√	√	
8	2.447	√	√	
9	2.452	√	√	
10	2.457	√	√	
11	2.462	√	√	
12	2.467		√	
13	2.472		√	
14	2.484			√

Table 4.1: *DSSS channels by geographical region*

fined, while in most of Europe 13 different channels have been specified. These channel definitions and their respective operating frequencies are summarized in Table 4.1.

The DSSS spreading mechanism is illustrated in Figure 4.3. In the case of 1 Mbps and 2 Mbps operation, each data bit of the incoming data sequence is combined with the 11-chip Barker spreading code by applying a binary adder (modulo 2). The resulting spreaded sequence has a higher bandwidth than the original data sequence. The factor of increase is 11 in this case and is referred to as *processing gain*. The processing gain improves the robustness of the transmitted signal.

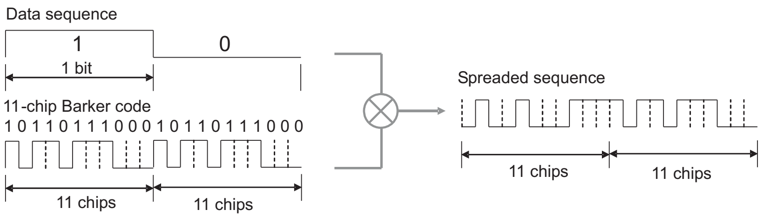

Figure 4.3: *Direct Sequence Spread Spectrum modulation*

The resulting spreaded sequence is used as the input for DBPSK or DQPSK modulator. These two modulation techniques are explained in the following subsection. Each DSSS channel occupies 22 MHz of bandwidth and has a spectral shape of a filtered $sin(x)/x$ function as shown in Figure 4.4.

Therefore, all channels that are spaces less than 22 MHz apart from each other overlap and cause inter-channel interference if used in close vicinity. Most of the channels, however, are only 5 MHz apart and do overlap. As a consequence, the DSSS channel definition for North America and most of Europe does only allow 3 non-overlapping channels, as shown in Figure 4.5. This restriction might cause problems in the case

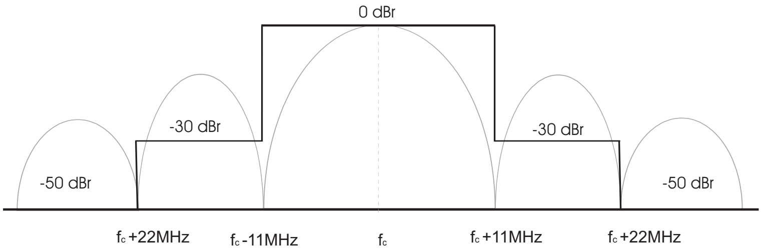

Figure 4.4: *DSSS transmit channel shape*

where large Wireless LAN systems are deployed. Three non-overlapping channels are not enough to cover a large area while preventing overlapping cells. Therefore, inter-channel interference occurs and lead to a performance degradation. Later chapters cover this topic in greater detail.

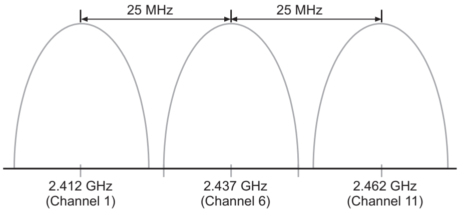

Figure 4.5: *Non-overlapping channels*

Phase shift keying (DBPSK, DQPSK)

The initial IEEE 802.11 standard defines two DSSS modulation techniques, Differential Binary Phase Shift Keying (DBPSK) and Differential

Quadrature Phase Shift Keying (DQPSK). Both mechanisms are special types of Phase Shift Keying, where the phase of the carrier frequency is varied in order to represent different binary symbols, and differential precoding and differential demodulation is applied.

In DBPSK two different signals are used to transmit the binary values 1 and 0. A maximum of 11 Mcps is possible. Each single data bit is represented by 11 chips. Thus, a data rate of 1 Mbps is supported. The concept of DBPSK is displayed in Figure 4.6.

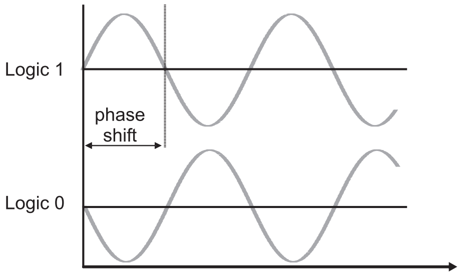

Figure 4.6: *Differential Binary Phase Shift Keying*

In the case of DQPSK, the input of the modulator is a combination of 2 bits (00, 01, 10, 11). The concept is shown in Figure 4.7. Four different transmitted phases are used. Each of these two-bit symbols is sent at 1 Mbps, such that an overall binary data rate of 2 Mbps is reached. The four-level modulation technique doubles the data rate while maintaining the same baud rate as the 1 Mbps signal.

The transmit power level for DSSS is restricted to 1 watt in North America and 100 milliwatts in Europe. For power values greater than 100 milliwatts, a power control mechanism should be implemented, that allows to specify a lower transmit power. A minimum transmit power of 1 milliwatts has been specified as well.

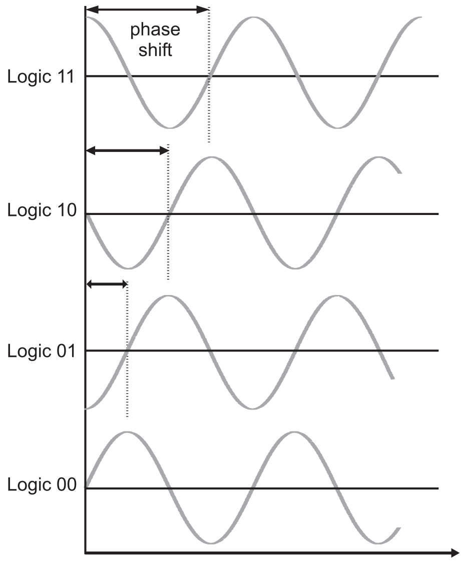

Figure 4.7: *Differential Quadrature Phase Shift Keying*

It is important to notice, that the entire header information of a physical layer packet (24 Bytes) shall always be transmitted using DBPSK (1 Mbps). This ensures the downward compatibility of devices that only support the DBPSK modulation. However, this also increases the overhead of the PHY layer, and thus decrease the maximum system performance. In the case where 2 Mbps are used, the increased overhead only slightly influences the system performance, but the higher the data rate, the larger the overhead gets. This is especially important in the case of higher data rates, such as 11 Mbps DSSS modulation. In later chapters, this matter is discussed in detail.

Complementary Code Keying (CCK)

The *Higher Speed Physical Layer Extension in the 2.4 GHz Band*, also known as the IEEE 802.11b standard, defines two more modulation techniques for the 2.4 GHz band in order to provide data rates of up to 11 Mbps. The 8-chip Complementary Code Keying (CCK) is specified as the default modulation type for the 5.5 Mbps and 11 Mbps data rates. However, in addition Packet Binary Convolutional Coding was included as an optional modulation technique and is explained later.

In contrast to the DBPSK and DQPSK modulation techniques described above, Complementary Code Keying (CCK) uses different complex spreading codes for modulation of the data bits. Each spreading code has a length of 8 chips. The chip rate is still 11 Mcps. The following formula is used to derive the CCK code words in order to spread both 5.5 Mbps and 11 Mbps.

$$\begin{aligned} C &= \{c_0, c_1, \cdots, c_7\} \\ &= e^{j(\varphi_1+\varphi_2+\varphi_3+\varphi_4)}, e^{j(\varphi_1+\varphi_3+\varphi_4)}, e^{j(\varphi_1+\varphi_2+\varphi_4)}, -e^{j(\varphi_1+\varphi_4)}, \\ &\quad e^{j(\varphi_1+\varphi_2+\varphi_3)}, e^{j(\varphi_1+\varphi_3)}, -e^{j(\varphi_1+\varphi_2)}, e^{j(\varphi_1)}\} \end{aligned}$$

C is the code word and consists of the eight complex chips $c_0, c_1, \cdots, c_7$. The code words are computed depending on a number of data bits. The term φ_1 is chosen from Table 4.2 depending on the first two data bits (d_0, d_1). It defines the phase change compared to the preceding phase φ_1 based on DQPSK. Therefore, the symbols of a data packet are numbered starting with "0" to determine odd and even symbols. The fourth and seventh chip (c_3 and c_6) are rotated 180 degrees to optimize the sequence correlation properties.

The terms $\varphi_2, \varphi_3, \varphi_4$ are chosen depending on the current data rate. In the case of 5.5 Mbps, 4 data bits are transmitted per symbol. The data bits d_2 and d_3 are used as follows.

(d_0, d_1)	phase change (even symbols)	phase change (odd symbols)
00	0	π
01	$\frac{\pi}{2}$	$\frac{3\pi}{2}$
10	π	0
11	$\frac{3\pi}{2}$	$\frac{\pi}{2}$

Table 4.2: *DQPSK encoding table*

$$\varphi_2 \quad = \quad d_2 \cdot \pi + \frac{\pi}{2}, \quad \varphi_3 = 0, \quad and \quad \varphi_4 = d_3 \cdot \pi.$$

Thus, the data bits d_2 and d_3 encode the basic symbol as specified in Table 4.3. It shows the value of the two data bits, and the appropriate complex chip values $c_0, \cdots, c_7$ which define the complex code words, the spreading code. The two input bits (d_0, d_1) only define the phase change that are used when transmitting the spreading code.

d_2, d_3	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7
00	1j	1	1j	-1	1j	1	-1j	1
01	-1j	-1	-1j	1	1j	1	-1j	1
10	-1j	1	-1j	-1	-1j	1	1j	1
11	1j	-1	1j	1	-1j	1	1j	1

Table 4.3: *5.5 Mbps CCK encoding table*

In the 11 Mbps mode, 8 bits are transmitted per symbol. The chip rate is still 11 Mcps. Again, the first two data bits (d_0 and d_1) are used to encode φ_1 as in the 5.5 Mbps case (see Table 4.2). The second pair of data bits (d_2 and d_3) is then used to encode φ_2, the third pair (d_4 and

d_5) encodes φ_3, and the fourth data bit pair (d_6 and d_7) encodes φ_4 as shown in Table 4.4.

data bits (d_i, d_{i+1})	phase
00	0
01	$\frac{\pi}{2}$
10	π
11	$\frac{3\pi}{2}$

Table 4.4: *11 Mbps CCK phase depending on data bits d_i and d_{i+1}*

Complementary Code Keying was developed by Intersil Inc. It is the most common modulation technique in today's Wireless LAN equipment. Texas Instruments developed a different modulation technique. It is called Packet Based Convolutional Coding (PBCC) and was included as an optional modulation scheme to the standard IEEE 802.11b. PBCC is described in the following.

Packet Binary Convolutional Coding (PBCC)

The Packet Binary Convolutional Coding (PBCC) scheme is an optional modulation technique for Wireless LAN in the 2.4 GHz band. It is defined in the IEEE 802.11b standard and allows data rates of 5.5 Mbps and 11 Mbps. Initially, PBCC was developed by Texas Instruments. It uses binary convolutional coding with a 64-state binary convolutional code (BCC) and a cover sequence.

The encoder is displayed in Figure 4.8. The input data is first encoded using a binary convolutional code (BCC) with rate 1/2. The output of the BCC encoder is mapped to a signal constellation using either 5.5 Mbps (BPSK) or 11 Mbps (QPSK) rate. In the 5.5 Mbps mode, the output pair of the BCC is taken serially to produce two BPSK symbols.

In the 11 Mbps mode, the output pair of the BCC produces one QPSK symbol. This gives a throughput of $\frac{1}{2}$ bit per symbol in the BPSK mode and 1 bit per symbol in the QPSK mode.

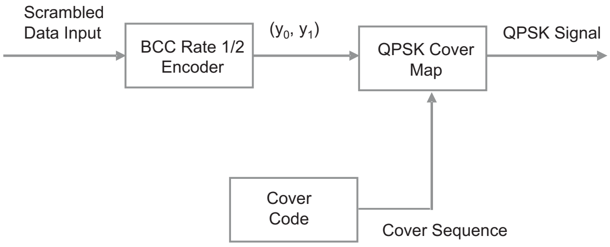

Figure 4.8: *PBCC modulator scheme*

The generator matrix of the binary convolutional code is given as

$$G = \left[D^6 + D^4 + D^3 + D + 1, D^6 + D^5 + D^4 + D^3 + D^2 + 1\right].$$

The block diagram of the resulting encoder is given in Figure 4.9. It consists of six memory elements and produces two output bits (y_0, y_1) for each input bit according to the generator matrix G. PBCC is packet based, which means that the encoder is set to state zero (i.e. all memory elements are initialized with the value zero at the beginning of each packet). At the end of a packet transmission, the encoder has to be placed in a known state as well in order to prevent that the last packet bits are less reliable. Therefore, at least six deterministic bits must be input immediately following the last data bit input to the convolutional encoder.

Finally, a pseudo-random cover sequence is generated. It is used to map the code to the appropriate symbol. The cover sequence is generated from a seed sequence. The 16-bit seed sequence is 0011001110001011. It is used to generate the 256-bit pseudo random coder sequence by taking

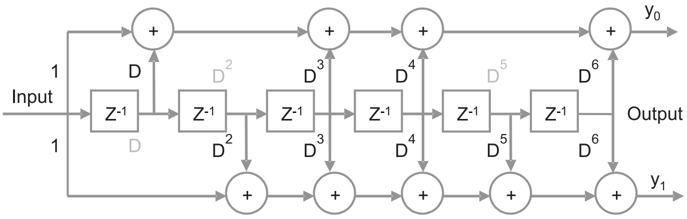

Figure 4.9: *Block diagram of the PBCC convolutional encoder*

the first sixteen bits of the sequence as the seed sequence, the second sixteen bits as the seed sequence cyclically left rotated by three, the third sixteen bits as the seed sequence cyclically left rotated by six, etc.

As of the writing of this work, PBCC is not supported by many devices on the market.

4.1.3 Orthogonal Frequency Division Multiplexing (OFDM)

Orthogonal Frequency Division Multiplexing (OFDM) is used in two different Wireless LAN standards: the IEEE 802.11g *Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band* [IEE03c] and the IEEE 802.11a *High-speed Physical Layer in the 5 GHz Band* [IEE99c]. In the 802.11g standard, OFDM is one of the possible extensions of IEEE 802.11b to achieve data rates of up to 54 Mbps in the 2.4 GHz band. It also defines further modulation techniques, which are explained at the end of this section. The IEEE 802.11a standard, on the other hand, defines OFDM as the main technology for Wireless LANs in the 5 GHz band. In the following the basic OFDM technology is explained as it is used in the IEEE 802.11a standard. Later, the OFDM extensions for the 2.4 GHz band as specified in the IEEE 802.11g standard are discussed.

OFDM Basics

Orthogonal Frequency Division Multiplexing is a multiplexing technique that splits up the available channel into a number of orthogonal subcarriers with lower data rates. The subcarriers are used for simultaneous transmission. Due to the increased symbol duration for the lower rate parallel subcarriers, the system can introduce a guard time for each OFDM symbol. Therefore, the intersymbol interference caused by multipath fading is almost completely eliminated. The subcarriers are modulated using one of the modulation techniques BPSK, QPSK, 16-QAM, or 64-QAM.

Frequency band	Channel number	Center Frequency (GHz)	Maximum Output Power (mW)
U-NNI lower band	36	5.180	40
	40	5.200	
	44	5.220	
	48	5.240	
U-NNI middle band	52	5.260	200
	56	5.280	
	60	5.300	
	64	5.320	
U-NNI upper band	149	5.745	800
	153	5.765	
	157	5.785	
	161	5.805	

Table 4.5: *IEEE 802.11a channels in the United States*

The IEEE 802.11a standard defines 12 different channels for use in the 5 GHz band. The original standard is dated 1999. It only specifies the channel settings for the United States as shown in Table 4.5. In the meantime, other regulatory domains were added. In Europe, the lower eight channels can be used. In the 5 GHz band, the channels numbering was chosen in a way that the whole area ranging from 5 GHz to 6 GHz is split into channels that are 5 MHz apart. Therefore, the lowest channel number for Wireless LAN in the 5 GHz band is 36.

All channels chosen for use in the 802.11a standard are 20 MHz apart. Therefore, they do not overlap as in the 2.4 GHz band. All eight channels are non-overlapping. For each of these channels, OFDM is used. It defines 52 subcarriers each 0.3126 MHz apart. 48 of these subcarriers are used for data transmission and 4 are pilot channels. If d_i are the complex QAM symbols, N_s is the number of subcarriers, T the symbol duration, and f_c the carrier frequency, then one OFDM symbol starting at $t = t_s$ can be written as

$$\begin{aligned} s(t) &= Re\left\{ \sum_{i=-\frac{N_s}{2}}^{\frac{N_s}{2}-1} d_{i+\frac{N_s}{2}}\, exp(j2\pi(f_c - \frac{i+0.5}{T})(t-t_s)) \right\}, \\ & \quad t_s \leq t \leq t_s + T \\ s(t) &= 0, t < t_s \quad \wedge \quad t > t_s + T. \end{aligned}$$

It can be shown mathematically that the calculation of the OFDM symbols is equivalent to the Inverse Fast Fourier Transform (IFFT), while the demodulation can be performed by the reverse operation, the Fast Fourier Transform. These two operations are almost identical, such that the same hardware can be used for both transmitter and receiver. Figure 4.10 shows the block diagram of an OFDM modem. The upper path is the transmitter chain, and the lower path corresponds to the receiver chain.

The subcarriers of a single OFDM signal can be shown to be orthogonal, i.e. all the subcarriers have an integer number of cycles in the time interval T. Therefore, intercarrier interference is avoided. Coding rates of 1/2 or 3/4 can be used and four different modulation techniques can be

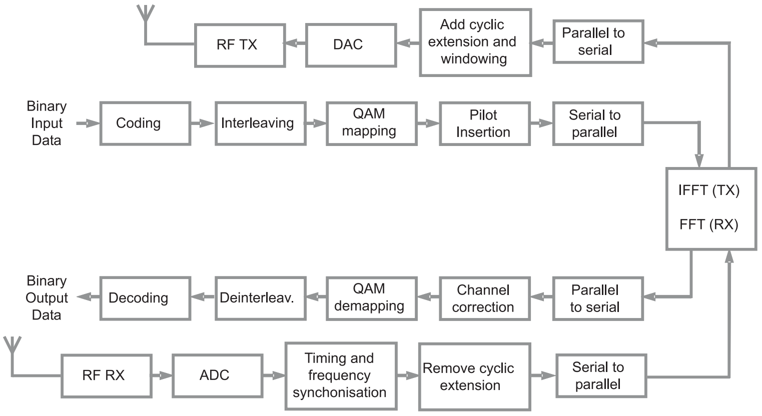

Figure 4.10: *Block diagram of the OFDM transceiver*

applied. BPSK and QPSK have been described above. Alternatively, 16-QAM and 64-QAM can be used to reach higher data rates. Figure 4.11 shows the rectangular constellations of Quadrature Phase Shift Keying (QPSK), 16-Quadrature Amplitude Modulation (QAM), and 64-QAM. The various modulation techniques allow to transmit different numbers of bits within one single modulated symbol. However, the 16-QAM and 64-QAM modulation techniques are less robust. In order to achieve the same bit error rate a better reception is necessary.

The different coding rates together with the four available modulation techniques allow to support a large set of data rates in IEEE 802.11a Wireless LANs. The standard specifies 6, 9, 12, 18, 24, 36, 48, and 54 Mbps. Mandatory data rates are 6, 12, and 24 Mbps. Table 4.6 shows the different scenarios.

The theoretical maximum data rate of 54 Mbps is sufficient for many practical environments and applications. Nevertheless, the main problem

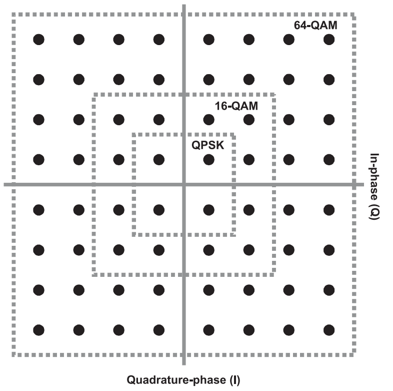

Figure 4.11: *QPSK, 16-QAM, and 64-QAM constellation*

of the IEEE 802.11a standard is the 5 GHz band. Such high frequencies are vulnerable to signal decay due to fading. This is especially important in indoor environments. Therefore, a high density of access points is necessary to reach an appropriate indoor coverage. The 2.4 GHz band is less susceptible to fading. The IEEE 802.11g standard extends the Wireless LAN standards for the 2.4 GHz band and also allows data rates of up to 54 Mbps. Thus, from the perspective of coverage and robustness, it is more appropriate for the use in indoor environments.

Hybrid Spread Spectrum coding method

The IEEE 802.11g *Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band* standard defines additional mandatory and optional modulation techniques for Wireless LAN in the 2.4 GHz band. The main goals of the standardization were the use of OFDM in the 2.4 GHz band and the downward compatibility.

Coding rate	Modulation	Bits per OFDM symbol	Data rate (Mbps)
1/2	BPSK	24	6
3/4	BPSK	36	9
1/2	QPSK	48	12
3/4	QPSK	72	18
1/2	16-QAM	96	18
3/4	16-QAM	144	24
1/2	64-QAM	192	36
3/4	64-QAM	216	54

Table 4.6: *IEEE 802.11a OFDM modulation techniques*

Pure OFDM operation as explained for the IEEE 802.11a standard is a mandatory modulation technique. It supports the same data rates, i.e. 6, 9, 12, 18, 24, 36, 48, and 54 Mbps. The preamble, header and data are transmitted using OFDM at the same data rate. However, it can not support the downward compatibility. Devices that only support IEEE 802.11b can not decode the packet header. On the other hand, the protocol overhead caused by the DBPSK modulated transmission of the preamble and header information is eliminated.

In order to achieve downward compatibility, the hybrid spread spectrum coding method DSSS-OFDM was added to the standard. The preamble and header information is transmitted using DBPSK with a data rate of 1 Mbps. Only the data portion is transmitted using OFDM with the higher data rates. Equipment that follows the IEEE 802.11b standard cannot decode the data portion of the transmission, but the preamble and header can be received and decoded. Therefore, devices using BPSK, QPSK, CCK, and DSSS-OFDM are interoperable in terms of the Carrier Sense Multiple Access protocol as explained in the next chapter.

As an option, the Packet Based Convolutional Coding (PBCC) supporting 22 and 33 Mbps was included in the IEEE 802.11g standard. This is just an extension of the optional PBCC modulation in the IEEE 802.11b standard.

The IEEE 802.11b and 802.11g standards support a rich set of different data rates, while the 2.4 GHz frequency band is less susceptible to fading than the 5 GHz band. Therefore, these two are the most important standards for the analysis of Wireless LAN in large indoor environments. Thus, the following chapters focus on the 2.4 GHz band.

4.2 Wireless Channel Models

The wireless medium is influenced by a number of different parameters. They all have a great effect on the quality of the performed transmissions in terms of bit error rate or packet error rate. Therefore, these factors have to be considered when analyzing the performance of a wireless system. In the following, the most important properties of a wireless channel are explained. For more information see [OP99], [PP01], and [Gas02].

The first important parameter to consider is the transmit power of the wireless device. It is measured in milliwatts. However, sometimes it is also referred to in terms of dBm, i.e. decibels referenced to 1 milliwatt. Different maximum transmit power values are specified in the standards, as explained in previous sections. The IEEE 802.11b standard, for example, defines a maximum of 100 milliwatts which is interchangeable with dBm as

$$\begin{aligned} 10 \cdot \log_{10}\left(\frac{100mW}{1mW}\right) &= 20dBm \\ 10^{\frac{20dBm}{10}} &= \frac{100mW}{1mW} \end{aligned}$$

In the following, if X defines a variable in decibels, then $\hat{X} = 10^{\frac{X}{10}}$ denotes the linear value.

However, the transmit power is measured as the effective radiated power (ERP) of the device. It is directly influenced by the antenna that is applied. Antennas change the focus of the radiated power. The power is concentrated in a certain area as specified by the radiation pattern of the antenna, which is given in horizontal and vertical direction. Figure 4.12 shows the radiation patterns for a 2 dBi standard dipole (rubber duck) antenna. The unit dBi describes the antenna gain in decibels referenced to an isotropic radiator. The torus on the left side indicates the 3-dimensional radiation pattern of the antenna. In technical references usually only the horizontal and vertical antenna patterns are shown.

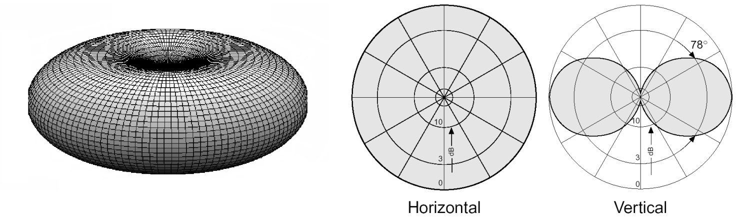

Figure 4.12: *Antenna pattern: horizontal and vertical radiation diagrams*

These horizontal and vertical antenna patterns show the additional gain that is reached in certain areas surrounding the antenna. The gain, however, has to be taken into account when calculating the maximum effective radiated power of a wireless device. Therefore, when high gain antennas are used, the actual transmit power of the device has to be lowered to keep the ERP below the allowed level.

On the way from the transmitter to the receiver, the transmitted signals are influenced by the effects of fading. *Large-scale Fading* represents the average signal power attenuation or the path loss due to motion over larger areas. *Small-scale Fading* refers to the dramatic changes in signal amplitude and phase due to small changes in the spatial separation between a receiver and transmitter.

Large-scale Fading or Path Loss defines the mean signal attenuation depending on the distance between sender and receiver, and is expressed in decibels. However, it greatly depends on the considered environment. Famous path loss models for free space propagation were developed by Okumura [Oku68] and Hata [Hat80]. However, these models account for factors such as antenna height, temperature, or carrier frequency. Therefore, many researchers use a simplified formula, that only accounts for the distance d from the transmitter. The path loss $L_p(d)$ is expressed in terms of the path loss $L_p(d_0)$ to the reference point at distance d_0 plus the additional free space loss depending on the actual distance d plus a random variable X_σ, the standard deviation of the path loss.

$$L_p(d)(dB) \quad = \quad L_s(d_0)(dB) + 10n\log_{10}\left(\frac{d}{d_0}\right) + X_\sigma(dB)$$

Here, n denotes the path loss exponent, X_σ defines a zero-mean, Gaussian random variable (in decibels) with standard deviation σ.

Small-scale Fading is itself influenced by different factors. However, measurements have shown that there are two main categories of small-scale fading. In the first category, the received signal is made up of multiple reflective rays plus a significant line-of-sight component. This type of fading is referred to as *Rician fading*, since the small-scale fading follows a Rician probability density function (pdf). The second category, where the line-of-sight link approaches an amplitude of zero, the small-scale fading follows a Rayleigh pdf. Therefore, it is called *Rayleigh fading*.

Knowing the transmit power and the fading for all the received signals at a receiver, the received signal power can be calculated. Let T_i be the transmit power of source i, $L_p^{i,j}$ the path loss from sender i to receiver j, $L_f^{i,j}$ the attenuation due to fast fading from sender i to receiver j. N denotes the thermal noise. The total interference from all other stations I at the receiving station i is calculated as

$$\hat{I}_i = \left(\frac{\hat{T}_k}{\sum_{k \neq i} \hat{L}_p^{k,j} \cdot \hat{L}_f^{k,j} + \hat{N}} \right)^{-1}$$

Then, the signal-to-noise ratio (SNR) for the signal transmitted from node i to node j in decibels can be expressed as

$$\mathrm{SNR}^{i,j} \quad = \quad T_i - L_p^{i,j} - L_f^{i,j} - I_i.$$

In the case of directional antennas, the antenna gain has to be taken into account, as well. For reasons of simplicity, the antenna gain was not included in the formula. The SNR can now be easily converted into the energy per transmitted bit by applying the processing gain P_g as in

$$\left(\frac{E_b}{N_0} \right)^{i,j} = \mathrm{SNR}^{i,j} + P_g^{i,j}.$$

The bit energy $\frac{E_b}{N_0}$ can then be used to calculate the bit error probability depending on the underlying modulation technique. In the case of Differential Binary and Quadrature Phase Shift Keying it is

$$\begin{aligned} \mathrm{BER}^{DBPSK} &= Q\left(\sqrt{S\hat{N}R \cdot \hat{P}_g^{DBPSK}} \right) = Q\left(\sqrt{\left(\frac{\hat{E}_b}{N_0} \right)} \right), \quad \text{and} \\ \mathrm{BER}^{DQPSK} &= Q\left(\sqrt{S\hat{N}R \cdot \hat{P}_g^{DQPSK}} \right) = Q\left(\sqrt{\left(\frac{\hat{E}_b}{N_0} \right)} \right), \end{aligned}$$

where the processing gain for DBPSK P_g^{DBPSK} is 11 and for DQPSK P_g^{DQPSK} it is 5.5 (see [LSDS01], [CSL02]). $Q(x)$ is defined as the area under the tail of the Gaussian probability density function with zero mean and unit variance

$$\begin{aligned} Q(x) &= \frac{1}{2}erfc\left(\frac{x}{\sqrt{2}}\right) \\ &= \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}dt. \end{aligned}$$

Due to the computational complexity of $Q(x)$, we use the following approximation

$$Q(x) = \frac{1}{2\pi}e^{-\frac{x^2}{2}} \cdot \left(\frac{8+9x^2+x^4}{15x+10x^3+x^5}\right).$$

In the case of CCK modulation, the IEEE uses a two-fold calculation for their analytical Wireless LAN model. First, the Symbol Error Rate (SER) is calculated. In the case for 5.5 Mbps CCK, the symbol error rate is given as

$$\mathrm{SER}_{5.5} = 15 \cdot Q(\sqrt{8 \cdot \hat{\mathrm{SNR}}}) + Q(\sqrt{16 \cdot \hat{\mathrm{SNR}}}).$$

As each symbol encodes 4 bits, the average BER is

$$\mathrm{BER}^{CCK_{5.5}} = \left(\frac{2^{4-1}}{2^4-1}\right)\mathrm{SER}_{5.5}.$$

And in the case of 11 Mbps CCK modulation, the symbol error rate is given as

$$\begin{aligned} \mathrm{SER}_{11} &= 24 \cdot Q(\sqrt{4 \cdot \hat{\mathrm{SNR}}}) + 16 \cdot Q(\sqrt{6 \cdot \hat{\mathrm{SNR}}}) \\ &\quad +174 \cdot Q(\sqrt{8 \cdot \hat{\mathrm{SNR}}}) + 16 \cdot Q(\sqrt{10 \cdot \hat{\mathrm{SNR}}}) \\ &\quad +24 \cdot Q(\sqrt{12 \cdot \hat{\mathrm{SNR}}}) + \cdot Q(\sqrt{16 \cdot \hat{\mathrm{SNR}}}). \end{aligned}$$

Each symbol encodes 8 bits. Thus, the average BER is

$$\mathrm{BER}^{CCK_{11}} = \left(\frac{2^{8-1}}{2^8-1}\right)\mathrm{SER}_{11}.$$

These formulas allow to calculate the BERs experienced by Wireless LAN clients in different environments. The settings that were chosen in this work are summarized in the following section.

4.3 Simulation settings

Our research focus lies on the performance capabilities of current and future Wireless LAN MAC protocols. The considered physical layer (PHY) is not the crucial factor. Therefore, only IEEE 802.11b Wireless LAN networks were considered. As explained above, these networks support four modulation techniques: DBPSK, DQPSK, CCK, and PBCC. The mandatory modulation techniques DBPSK, DQPSK, and CCK are taken into account in our simulations, while the optional PBCC modulation technique is ignored.

We choose the wireless channel parameters in accordance with the IEEE recommendations for the analytical and simulation models for the BER of IEEE 802.11b transmissions. In the following these settings are summarized.

The transmit power is set to 100 mW or 20 dBm. This is the maximum allowed transmit power for IEEE 802.11b devices in Europe. Omnidirectional antenna patterns are assumed. Therefore, the antenna exhibits no additional gain in terms of transmit or received power.

The path loss model is a slight modification of the free space propagation model.

$$\begin{aligned} L_p(d) &= 40.2 + 20 \cdot \log(d), \quad d < 8m, \quad \text{and} \\ L_p(d) &= 58.5 + 33 \cdot \log\left(\frac{d}{8}\right), \quad d \geq 8m \end{aligned}$$

However, the model is not valid for distances smaller than about 0.5m due to near-field and other implementation effects. Small scale fading is not considered in our simulations.

The formulas for the calculation of the bit error rates (BER) were taken as explained in the previous section.

5 LLC - WLAN Logical Link Control

> If you can't get rid of the skeleton in your closet, you'd best teach it to dance. George Bernard Shaw (1856-1950)

The initial Wireless LAN standard IEEE 802.11 [IEE99a] defines a cost-effective technology to allow an easy installation of a network in places where, for example, a wired network is not an option. In this 1999 version, data rates were restricted to 1 and 2 Mbps. In the meantime, many extensions to the Wireless LAN standard have been defined, as already discussed in Chapter 2.

This chapter starts with an in-depth explanation of the different access mechanisms. This includes the basic *Carrier-Sense Multiple Access with Collision Avoidance* (CSMA/CA) protocol as well as the Quality of Service (QoS) enabling Medium Access Control (MAC) protocols of polling and prioritization. The second part of this chapter discusses the performance issues that arise in various scenarios when these different MAC protocols are utilized. The goal is to evaluate their capabilities and whether they meet future 4G requirements. Finally, various WLAN handover mechanisms on layer two of the ISO/OSI reference model are summarized and evaluated as well.

5.1 Architecture of IEEE 802.11 Wireless LANs

A Wireless LAN network can be configured in two different modes of operation. The *ad-hoc mode* allows all involved clients to directly communicate with each other as long as they are in reception range. They form an Independent Basic Service Set (IBSS). In the *infrastructure mode*, on the other hand, a special station, the access point, is responsible for the routing of all traffic. Stations cannot communicate directly, but they always have to communicate with the access point, which forwards the data packets to the receiving station. A single access point and all the stations that communicate through this access point form a Basic Service Set (BSS). Several BSSs can be connected via a Distribution System (DS). The main function of the access point is to distribute the data packets between the various stations within its BSS and to forward the data to the Distribution System (DS). The combination of several BSSs that are connected through a DS is referred to as an Extended Service Set (ESS). The different topologies are displayed in Figure 5.1.

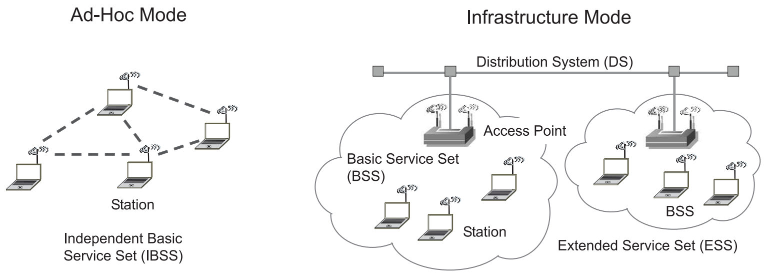

Figure 5.1: *Ad-hoc versus Infrastructure Mode*

We want to study the capabilities of Wireless LAN as an extension of legacy mobile networks in the context of upcoming 4G systems. The ad-hoc mode is assumed not to play an important role in such a scenario. Therefore, we solely focus on the infrastructure mode in the remainder of this work.

The Wireless LAN standard IEEE 802.11 and all of its extensions define the two lowest layers of the ISO/OSI layer model, the Physical layer (PHY) and the Logical Link Control layer (LLC). The PHY layer was already described in Chapter 4. In this chapter, the WLAN specifications on the LLC layer are studied.

5.2 Medium Access Control Protocol

The main part of the LLC are the different Medium Access Control (MAC) mechanisms which define access control to the shared wireless medium. The original IEEE 802.11 standard defined two different access mechanisms, which can coexist in a Wireless LAN network. The distributed access protocol CSMA/CA and a centralized polling approach. These mechanisms and their extensions as defined for example in the IEEE 802.11e standard ([IEE03d], [IEE03e]) are discussed in the following.

5.2.1 Definitions

Wireless LAN is the wireless counterpart of Ethernet. Both technologies focus on local area networks. Therefore, these two have a lot in common. However, the wireless medium causes additional problems, which have to be taken care of by the Wireless LAN MAC protocol (see [PP01], [OP99], [Gei01], [Gas02]).

The best known problem is that of *Hidden Nodes*. Two WLAN stations are defined to be in a Hidden Node relation, if they are not in the reception range of each other, but there is a destination station that both have in common. A typical case of two Hidden Nodes is shown in Figure 5.2. Two stations are associated with the same access point. However, they are located at opposite directions, and thus are not in the reception range of each other as indicated by the two circles surrounding the individual client stations. Both of these stations, however, want to transmit data packets to the access point.

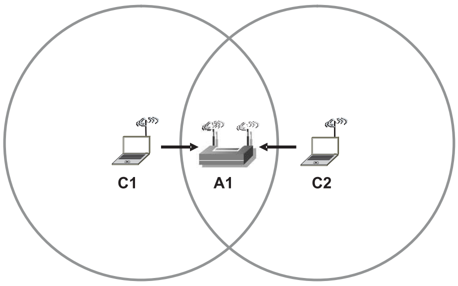

Figure 5.2: *Hidden Nodes*

The carrier sensing mechanism in the two client stations is not sufficient in such a case. Concurrent access to the medium leads to a high probability of collisions at the access point. Thus, the Hidden Node problem leads to a performance degradation.

Another problem specific to the wireless medium is that of *Exposed Nodes* as shown in Figure 5.3. Two client stations (C2, C3) are in a *Exposed Node* relation if they are in the reception range of each other, but their individual destination stations are only in the reception range of the transmitting station. In this case, the transmissions of the two source stations do collide at the transmitting stations, but not at the receiving stations.

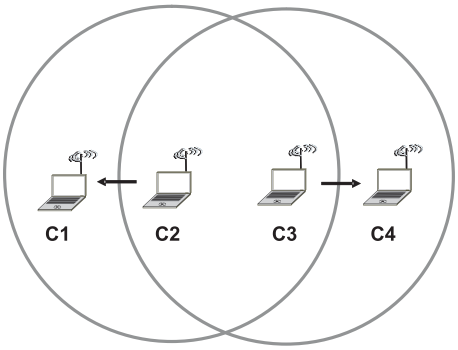

Figure 5.3: *Exposed Nodes*

In such a case, both "inner" stations can simultaneously transmit successfully. The overall bandwidth can be increased. However, these cases are hard to detect.

5.2.2 Overview of the MAC protocols

The Wireless LAN Medium Access Control mechanisms are summarized in Figure 5.4. The basic access technology is defined as the Distributed Coordination Function (DCF). It defines a distributed and contention-based access scheme and works as the basis for all MAC layer extensions. Section 5.2.3 explains the DCF protocol and discusses its properties.

Two QoS-enabling MAC extensions were defined on top of DCF. The Point Coordination Function (PCF) defines a polling mechanism which allows individual stations to get contention-free access to the wireless medium. However, Section 5.2.4 shows that this approach is not appropriate to support QoS levels necessary for Wireless LANs in 4G networks.

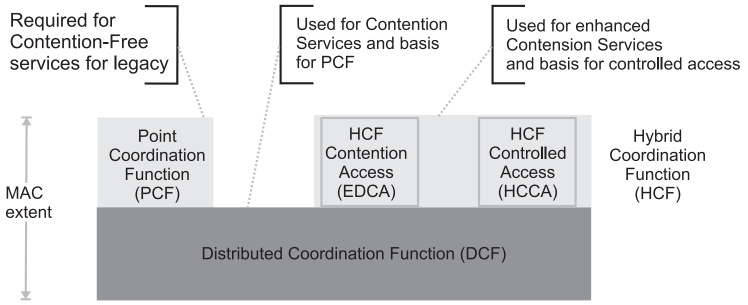

Figure 5.4: *Wireless LAN MAC layer architecture*

This has long been identified by the IEEE standardization bodies. Therefore, to enhance the QoS capabilities of WLAN in such environments, the IEEE defined the Hybrid Coordination Function (HCF). Similar to the PCF, the HCF is implemented on top of the DCF functionality. In contrast to the PCF, the HCF allows a highly sophisticated QoS support. It allows the simultaneous usage of applications with different QoS requirements, while still keeping the performance at an acceptable level. The HCF protocol is explained in Section 5.2.5.

5.2.3 Distributed Coordination Function

As explained in the former section, the Distributed Coordination Function defines the basic access scheme for Wireless LAN stations. It mainly consists of the Ethernet-like *Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA)* mechanism. All stations equally compete for access to the medium. The basic approach is shown in Figure 5.5. It depicts two clients that want to transmit a data packet to the access point.

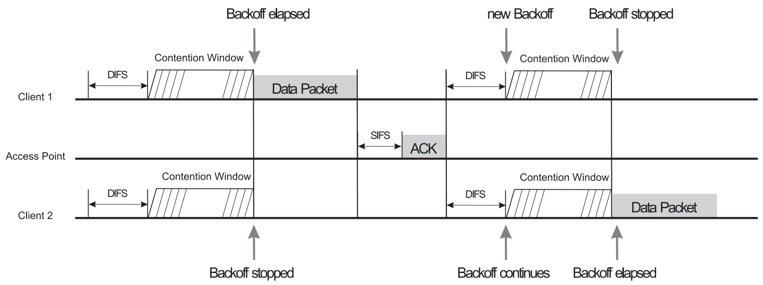

Figure 5.5: *Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA)*

Contention Window and Backoff

Before accessing the medium, the stations have to perform sense the carrier for a given amount of time, referred to as the Distributed (Coordination Function) Interframe Space (DIFS). If the medium is found idle, the clients are allowed to start their individual backoff algorithm. As for Ethernet, the clients choose a random number of time slots the medium has to be idle before the actual transmission can be started. In the case of Wireless LAN, the number of backoff slots is chosen uniformly distributed from the interval $[0, CW]$, where CW (Contention Window) defines the maximum Contention Window size. The minimum CW value is defined by the CW_{min} parameter, which is defined to be 31 for DSSS operation.

The main difference to Ethernet is that such a backoff has to be performed prior to any transmission. The Ethernet protocol states that a backoff has to be performed only after a collision.

After the backoff of a client has elapsed, the data transmission is started. All other involved clients within reception range find the channel busy. Therefore, they stop their backoff procedure and wait for the channel to become idle again for a period of DIFS before the backoff is continued.

Collision Avoidance and Error Recovery

The Ethernet protocol allows a client to recognize a collision on the medium. Each transmitting station simply compares the signal on the medium to the signal that it sends out. If the two are different, a collision is detected. However, Wireless LAN does not allow such an easy approach. On the one hand, cheap hardware was one of the most important design goals. Therefore, the chip sets only implement half-duplex operating modes. Either a station transmits or receives. Therefore, the Ethernet-like collision detection mechanism cannot be used with such client adapters. On the other hand, certain situations might arise in Wireless LAN environments, where a collision only occurs on the receiving side, for example in the case of Hidden Nodes. Therefore, different ways to detect and resolve collisions had to be found.

In the case of Wireless LAN, a simple acknowledgment scheme is used. It specifies that the receiving station has to acknowledge successful transmissions. Therefore, the access point in Figure 5.5 answers the reception of the data packet by replying with an ACK packet. If the client receives the ACK packet, it assumes a correct transmission. In any other case (either the data packet was not received due to a collision or the ACK packet was disturbed on the wireless medium), the data packet has to be retransmitted. A binary exponential backoff algorithm is employed to resolve lost packet conditions. The CW value is recalculated by $CW' = (CW + 1) \cdot 2 - 1$ before the random number of backoff slots is chosen. This operation is repeated in case of another collision. The CW value is upper bound by the CW_{max} value, defined as 1023 in the standard, which is reached for the 6th retransmission of a single packet.

Interframe Spaces

To send the ACK packet, a client also has to perform carrier sensing. However, in case of ACK packets or any other management frame, the Short Interframe Space (SIFS) is used instead of the DIFS. The reason is that the shorter SIFS interval assures that a station that wants to transmit an ACK packet gets access to the medium and not any other station that wants to transmit a data packet after a carrier sensing interval of DIFS. Two more interframe spaces are defined for Wireless LANs. Their relation is shown in Figure 5.6 and their purposes are explained below.

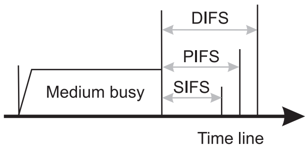

Figure 5.6: *Interframe Spacing relationship*

Short Interframe Space: The SIFS is used for the highest-priority transmissions, such as RTS/CTS frames and positive acknowledgment.

Point (Coordination Function) Interframe Space: The PIFS is used by the PCF during contention-free operation.

Distributed (Coordination Function) Interframe Space: The DIFS is the minimum medium idle time for contention-based services.

Extended Interframe Space: The EIFS is used only when there is an error in frame transmission.

Additional interframe spaces are defined for the Hybrid Coordination Function (HCF) operating modes and are introduced later in this chapter.

Carrier Sense mechanisms

Two different types of carrier sense mechanisms exist. The first one, as already mentioned above, is the physical carrier sense mechanism provided by the Physical layer. It scans the mediums for a signal. The second one is a virtual carrier sense mechanism. It utilizes the duration field that most Wireless LAN packets carry. These duration fields are used to indicate the duration of a transmission including the time for any other necessary packets, e.g. the ACK frame. A station reads the value of each received data packet and sets its Network Allocation Vector (NAV) to the specified amount of time. As long as the NAV counter of a station has not been count down to zero, the station is not allowed to start a data transmission even if the physical carrier sensing indicates an idle channel.

RTS/CTS mechanism

In a Hidden Node scenario, however, the carrier sensing mechanisms do not work properly. Consider a situation where two stations that are hidden from each other transmit to the same destination node, e.g. an access point. Such a case is shown in Figure 5.7. After the two stations perform their physical carrier sensing as well as the backoff operation, they start transmitting their data packets. Since they are hidden from each other, the physical carrier sensing does not recognize an ongoing transmission of a Hidden Node.

Therefore, the two stations both transmit simultaneously, which leads to an immediate collision of the packets at the access point, which in turn drops the corrupted data packets. Both stations wait for the ACK packet until a timeout elapses and start the retransmission operation. Depending on the size of the data packets, the collision probability can reach high levels leading to a drastic reduction of the available bandwidth.

To account for this problem, the Request-To-Send/Clear-To-Send

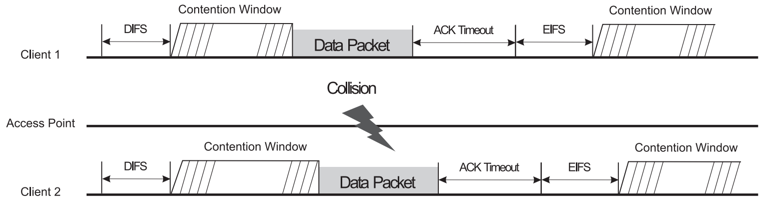

Figure 5.7: *pure CSMA/CA operation in Hidden Node cases*

mechanism was included in the Wireless LAN standard. It defines two short control frames, the Request-To-Send (RTS) and the Clear-To-Send (CTS) frame, which are exchanged by the sending and receiving station prior to the data transmission. Figure 5.8 shows that all clients within the reception range of either the sending or the receiving station receive at least one of the two frames and set their virtual carrier sensing timer NAV to the appropriate value. Here, the NAV is set to the time of the complete transmission including the RTS, CTS, data, and ACK frame.

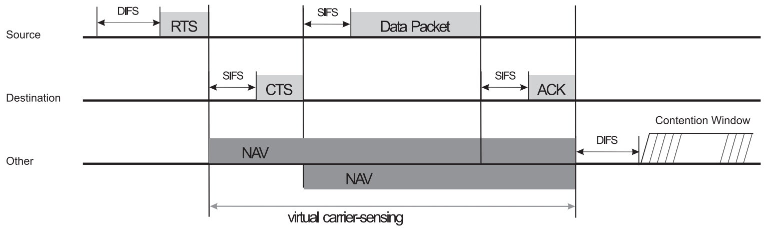

Figure 5.8: *RTS/CTS operation and virtual carrier sensing (NAV)*

The RTS and CTS control frames are very short compared to data packets, such that the collision probability in case of Hidden Nodes can be clearly reduced. Therefore, the performance of the system can be increased. On the downside, the additional exchange of the control frames adds to the overhead of the protocol, which leads to a performance degradation, especially in cases without Hidden Nodes where the RTS/CTS exchange actually is not necessary. This obvious trade-off is examined in Section 5.3.1. To keep the overhead at a low level, especially in cases with short data frames, the stations use an RTS/CTS threshold, which defines the minimum length of a data packet when RTS/CTS should be performed. In any other case, the RTS/CTS operation is omitted.

Fragmentation

The data portion of a Wireless LAN data packet has a maximum length of 2312 Bytes. However, higher layer data packets might exhibit a larger size. Such large data packets have to be broken into smaller pieces to fit the Wireless LAN packet structure. Fragmentation may also help to improve the reliability of the data transmission especially if the packet error probability is large, e.g. in cases with high values of interference.

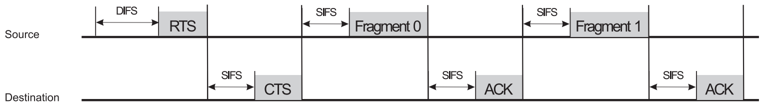

Figure 5.9: *Fragmentation*

Fragmentation takes place if the higher layer packet exceeds the fragmentation threshold. All fragments are assigned with a sequence number to help with reassembly. Since the fragmentation threshold is usually set to a larger value than the RTS/CTS threshold, the fragmentation procedure is usually initiated by RTS and CTS packets as shown in Figure 5.9. After RTS/CTS transmission, the station sends the various fragments subsequently, while utilizing the SIFS interframe space to receive the highest priority and to prevent non-synchronized stations from starting any transmission. Acknowledgments (ACKs) are sent by the receiving station for each individual fragment that was received. In case that a fragment is not correctly received, the receiving station drops the packet, while a timeout at the sending side indicates the lost fragment. The retransmission of the fragments is performed selectively, i.e. only the lost fragments are resent.

5.2.4 Point Coordination Function

The last section introduced the Distributed Coordination Function, an Ethernet-like access protocol, where all the stations content for the medium. Stations or applications can not be assigned different priorities on the basis of CSMA/CA. To allow the support for near real-time services within 802.11 environments, the PCF extension of DCF was included in the standard.

The Point Coordination Function (PCF) is based on DCF and implements a simple polling mechanism. It was designed to work in conjunction with the normal DCF mode. A special station, usually the access point, serves as the polling master or Point Coordinator (PC). It is the central organizational unit and is responsible for the organization of alternating contention-free and DCF-based service intervals. To do so, the PC regularly transmits Beacon frames, which indicate the start of a Contention-Free Period (CFP). In addition, the Beacon frame sets the NAV of all stations in reception range to the complete duration of the CFP. Therefore, DCF-based stations defer their transmission after the CFP.

DCF-based stations that could not receive the Beacon frame, and are, thus, not synchronized with the CFP, are kept from accessing the medium by the utilization of SIFS and PIFS interframe spaces. The PCF

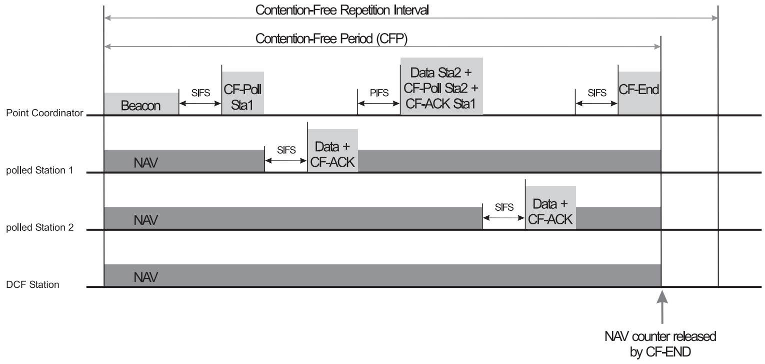

Figure 5.10: *Example of the Point Coordination Function (PCF) polling mechanism*

clients always gets priority over the longer DIFS interframe space used by the DCF clients.

An example of the PCF polling mechanism is given in Figure 5.10. The Point Coordinator initiates the CFP by transmitting a Beacon frame. All the stations set their initial NAV counters to CFPMaxDuration, which is set to 200 Kμsec by default. Kμsec defines Kilo microseconds. (One Kμsec equals 1,024 microseconds.)

Now, the CFP is in progress and the PC transmits a Data, CF-Poll, Data+CF-Poll, or CF-End frame after SIFS. A station receiving such a directed frame responds after SIFS. If a station receives a Poll frame, it may transmit a data frame to the PC. If the station has nothing to transmit, it acknowledges the poll frame by sending a null response frame in order to distinguish a no-traffic situation from a collision or otherwise lost packet. All of these frames can also be used to acknowledge the correct reception of the last packet, even if its sender is not the receiver of the frame, i.e. a CF-Poll+CF-ACK frame can be used to poll one station and to send an ACK to another station.

CF-End or CF-End+ACK frames mark the end of a CFP. This can be due to the elapsed CFP. However, the PC can also end a CFP if there is no more traffic to send. Therefore, no time is wasted, and the remaining time, until the next Beacon frame is scheduled, can be used for contention-based access. This notification is used by the clients to release their NAV counters. The contention-free service has to end no later than the maximum duration from the expected beginning point, which is referred to as the Target Beacon Transmission Time (TBTT).

The polling list

Within the CFP, the Point Coordinator polls the stations. Therefore, the PC keeps a polling list to keep track of these CF-Pollable stations. Only the PC is allowed to initiate a transmission, while all other stations have to remain idle. They have to wait until they receive a CF-Poll frame from the PC.

There is no defined policy on how to process the polling list. A straightforward approach surely is round-robin, but the standard does not define any restrictions. It only states, that at least one station has to be polled within a single CFP if there are stations on the polling list. And the polling list shall be processed in ascending order of a numeric ID of the stations.

If a data frame transmitted from a station is not acknowledged, the station does not retransmit the frame unless it is polled again. In contrast to DCF, no retransmission counter is used under PCF.

The polling list itself is only updated during the association or reassociation of a station. As discussed in Section 5.4.3, a station has to associate/reassociate with an access point before data frames can be exchanged. Information necessary for the communication is exchanged.

In these Association/Reassociation requests, the station can set its CF-Pollable subfield, which indicates to the access point the CF-Pollability of the station.

Arising problems with PCF

The Point Coordination Function was standardized to allow stations to use near real-time traffic ([CGKT02], [KW04], [SK96]). One main design goal was simplicity for easy and cheap implementation. However, this also leads to a number of problems of the PCF.

First of all, stations do not have any means to distinguish between PCF and DCF data traffic. They only maintain a single queue for all data traffic that has to be sent. Whenever a station is polled or sends a data packet in the contention-based period, it simply takes the first packet out of the queue. Therefore, PCF does not allow to support the simultaneous use of applications with different QoS requirements, such as Web and voice traffic. It can merely be used to distinguish between stations that have a "prioritized" access and those that do not. This could be used, for example, to support different Web clients and Voice clients. However, it certainly does not fit the needs of today's highly integrated voice and data terminals, i.e. Personal Digital Assistants (PDA).

Secondly, there is the *deferred Beacon problem of PCF* [VC03]. It states that the transmission of a Beacon frame can be delayed if the medium is not idle at the TBTT. If the medium is busy at TBTT, the PC waits for the medium to become idle for PIFS before Beacon transmission. It can easily be seen that this point in time is upper bound but not predictable. In such a case, the PC foreshortens the CFP by the time the Beacon frame was delayed.

Thirdly, the duration of a single poll procedure is not predictable, since a station is allowed to transmit a frame with arbitrary length, upper bound by the maximum allowed packet length of Wireless LAN. It cannot be foreseen when the next station is polled. Therefore, the service period that a single CF-pollable station receives, can not be predicted in advance, such that no fixed QoS guarantees can be kept.

Finally, problems arise in overlapping or co-located cells. Such overlaps in terms of coverage and channel frequency usually destroy the contention-free service completely. The Contention Free Periods of the involved access points experience an overlap in time, since no synchronization mechanism is implemented. In most cases the Contention Free Periods are set to 90 percent of the Contention-Free Repetition Interval, which even makes a potential synchronization impossible. However, the timely overlapping periods of polling, definitely cause regular collisions which are not taken care of by the MAC protocol. Therefore, PCF operation does not suffice the requirements for a Wireless LAN being integrated into a 4G environment and is, thus, not further analyzed in the remainder.

5.2.5 Hybrid Coordination Function (HCF)

QoS requiring applications of 4G especially include video, audio, real/time Voice over IP (VoIP), and other multimedia applications. These applications have different requirements regarding bandwidth, throughput, end-to-end delay, jitter, packet loss, or Mean Opinion Score (MOS). No service differentiation is provided by the IEEE 802.11 mechanisms as defined in [IEE99a] or [IEE99b]. The DCF and PCF mechanisms of IEEE 802.11 do not provide QoS to Wireless LAN stations. There is no service differentiation in DCF and PCF of traffic received from the Application layer. All frames received from the Application layer are enqueued in the same higher-layer data queue. This causes all data streams received from the Application layer to have the same (best-effort) priority access on the MAC layer.

The IEEE 802.11e working group is, therefore, working on a new

WLAN standard extension, which enhances the original MAC protocol to provide QoS and service differentiation on the MAC layer. Simple studies of the IEEE 802.11e standard can be found in [VCBS01] and [BVC01]. This section describes the MAC enhancements for QoS which comply to the IEEE 802.11e draft standard ([IEE03d], [IEE03e]). The final version of the standard has not been released yet, but it has reached a rather stable state, such that the results derived in later sections also apply to the final version.

Overview

The Wireless LAN QoS medium access is managed by the Hybrid Coordination Function (HCF). The HCF combines functionality from the DCF and PCF with some enhanced QoS-specific mechanisms for QoS frame transfers in a single MAC protocol. The HCF offers a contention-based and a contention-free access method to provide QoS stations (QSTA) with prioritized and parameterized QoS access to the wireless medium, while still supporting best-effort traffic to non-QoS STAs.

HCF supports a consistent set of frame formats and frame exchange sequences that QSTAs use during both the Contention Period (CP) and the Contention-Free Period (CFP). The contention-based service is defined as Enhanced Distributed Channel Access (EDCA) in contrast to the contention-free based service which is provided by the HCF Controlled Channel Access (HCCA). HCCA is the counterpart of the PCF mechanism in HCF and is also based on a polling mechanism. The goal of QSTAs is to obtain Transmission Opportunities (TXOP), which are the basic units of allocation of the right to transmit onto the wireless medium, using both EDCA and HCCA. A TXOP is defined by the start time and by a time interval which determines the maximum duration of the TXOP. If a TXOP is obtained using EDCA, it is called EDCA TXOP. If a TXOP is obtained using HCCA, it is called a polled TXOP.

htb

Priority	User Priority	802.1D Designation	Access Category	Designation (Informative)
Lowest	1	Background (BK)	0	Best Effort
	2	-	0	Best Effort
	0	Best Effort (BE)	0	Best Effort
	3	Excellent Effort (EE)	1	Video Probe
	4	Controlled Load (CL)	2	Video
	5	Video (VI)	2	Video
	6	Voice (VO)	3	Voice
Highest	7	Network Control (NC)	3	Voice

Table 5.1: *User Priority to Access Category mapping*

Enhanced Distributed Channel Access (EDCA)

The Enhanced Distributed Channel Access mechanism is the contention-based medium access mechanism for HCF. It is based on differentiating User Priorities (UP), as summarized in Table 5.1. These UPs define how the data traffic is to be delivered. According to the IEEE 802.1D standard [IEE98], they range from zero (0), the lowest priority designated to the transport of best-effort traffic, up to seven (7), the highest priority used for network control traffic. The IEEE 802.1D standard also assigns QoS parameters to the different UPs, e.g. the Video UP (5) is supposed to garantee delays of less than 100 ms. However, the EDCA mechanism can only differentiate between the four different categories, the Access Categories (AC). Therefore, the User Priorities as defined in the IEEE 802.1D standard are mapped to the four Access Categories shown in Table 5.1.

In order to achieve a service differentiation according to the ACs on the Wireless LAN MAC layer, three different parameters of the MAC protocol are utilized: (i) the amount of time a station performs carrier sensing before the backoff is initiated, the Interframe Space, (ii) the length of the Contention Window to be used for a backoff, and (iii) the duration a station may transmit after it acquires a TXOP.

All data traffic is mapped to ACs and inserted into one of the four transmit queues as illustrated in Figure 5.11. Each of the transmit queues is processed by an individual Channel Access Function (CAF), i.e. a virtual DCF stations with its own MAC layer parameters. The four virtual stations compete with each other for medium access. The Access Catogories are sorted from AC0 to AC3, with AC3 having the highest priority for medium access.

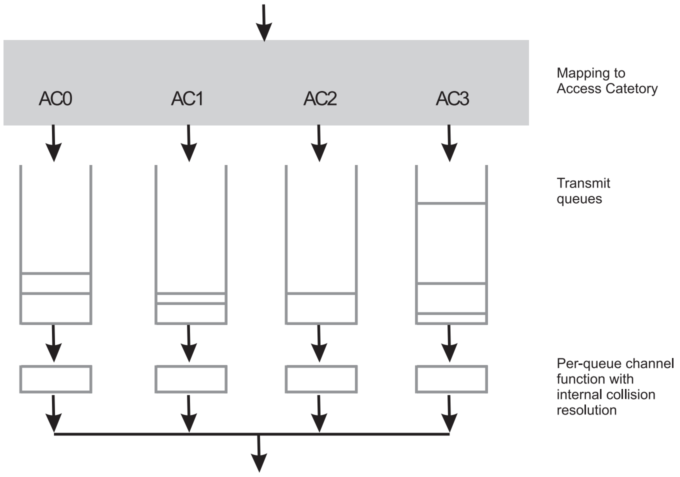

Figure 5.11: *Hybrid Coordination Function Access Category mappings*

The MAC layer parameters that are used by each of the virtual stations are summarized as EDCA Parameter Set in the standard and consist of Arbitration Interframe Space (AIFS), AIFS Number (AIFSN), CW_{min}, CW_{max}, and TXOPLimit. The specific values for these parameters are discussed later.

EDCA Transmission Opportunities (TXOP)

Each AC within one QSTA works like a virtual station. It contends for the channel access to the wireless medium and performs the backoff if necessary. Each transmit queue of an AC maintains its own Channel Access Function (CAF) and each CAF has its own Medium Occupancy Timer, which is used to qualify the validity of a TXOP for this transmit queue. An EDCA TXOP occurs, when EDCA rules permit access to the medium. The Medium Occupancy Timer is set to the TXOPLimit of the corresponding AC. Continuation of use of the wireless medium is granted, when the CAF retains medium access following the completion of a frame exchange sequence. The Medium Occupancy timer continues to count down to zero until the TXOPLimit has elapsed and is not reloaded for the new frame exchange sequence. This implies that the particular CAF can transmit several frames, if it has the TXOP and if the TXOPLimit has not been reached.

An internal collision occurs if two or more CAFs of a single STA obtain the TXOP at the same time. This is resolved internally in the QSTA. The CAF with the higher priority gets the TXOP, while the lower-prioritized CAF behaves as if there was an external collision on the wireless medium.

Obtaining an EDCA TXOP

To distinguish between ACs and backoff functions, different interframe spaces are introduced for each CAF. Instead of waiting DIFS before a transmission, the Arbitration Interframe Space (AIFS) is used. The duration of AIFS depends on the AC and is a duration derived from the value AIFSN[AC] by the relation $AIFS[AC] = AIFSN[AC] \cdot SlotTime + SIFS$ and is measured in seconds. The AIFS is at least DIFS (50 μsec) and can be enlarged individually for each AC. All IFS relations are shown in Figure 5.12.

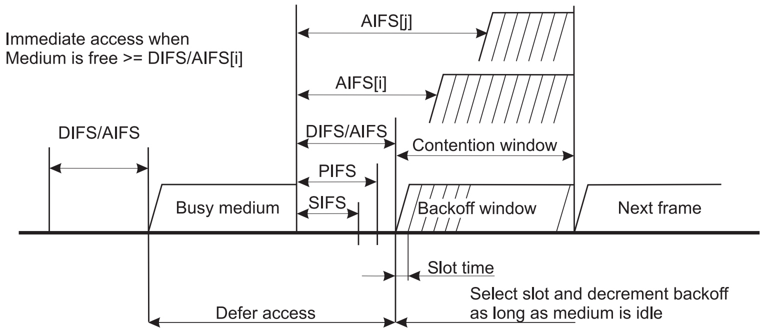

Figure 5.12: *EDCA Interframe Spaces*

Like in DCF, each CAF senses the medium to be idle for at least a minimum duration of AIFS[AC]. A transmission can begin immediately if the medium was sensed idle for this duration, the backoff timer for that CAF is zero, and these conditions are not simultaneously met by an AC of higher UP. Otherwise, the QSTA defers until the ongoing transmission has finished and the medium becomes idle. After deferral, each CAF senses the medium to be idle for AIFS[AC] again before continuing with the backoff procedure. If errors occurred during a previous frame exchange, each CAF waits for $EIFS - DIFS + AIFS[AC]$ before starting the backoff procedure.

EDCA Backoff Procedure

The backoff procedure consists of the backoff function like in DCF, which contains a value measured in backoff slots. Unlike DCF, in EDCA a single QSTA contains several backoff timers, one for each CAF. These multiple backoff timers run in parallel. The random backoff is defined by the Contention Window (CW) parameter calculated from the CW_{min} and CW_{max} parameters. To distinguish between ACs, each AC has its own CW parameter and different CW_{min} and CW_{max} values are used for each CAF as shown in Table 5.2. In most cases, an AC with higher priority is assigned a shorter CW to ensure that on average the higher-priority AC can transmit before the lower-priority AC. Each CW parameter is initialized with CW_{min} which is reset after each successful frame transmission.

AC	CW_{min}	CW_{max}	AIFSN
0	CW_{min}	CW_{max}	2
1	CW_{min}	CW_{max}	1
2	$(CW_{min} + 1)/2$-1	CW_{min}	1
3	$(CW_{min} + 1)/4$ -1	$(CW_{min} +1)/2$-1	1

Table 5.2: *EDCA Access Category parameters*

The backoff procedure for a CAF is invoked if an AC requests for a TXOP and the medium is busy as indicated by the physical or virtual carrier sensing mechanism. Each CAF also waits for a Contention Window after expiration of the EDCA TXOP if the TXOPLimit has been reached. The backoff procedure is also invoked if a transmission fails.

In case that a transmission fails, the CW value is incremented to $CW = (CW + 1) \cdot 2 - 1$ if $CW < CW_{max}$ and it is set to $CW = CW_{max}$, if the retry limit has been reached. This CW adaptation algorithm is also used under DCF. In contrast to the DCF operating mode, the random number of backoff slots is chosen uniformly distributed from the interval $[1, CW+1]$ instead of $[0, CW]$. Therefore, a backoff value of zero is not possible. The backoff timer is decremented at the end of each backoff slot if the medium has been idle for the duration of that slot.

TXOP Continuation

Continuation of an EDCA TXOP is granted to a CAF after waiting for SIFS and after successful completion of a frame exchange sequence. The continuation is only permitted for the transmission of a frame of the same AC that was granted the EDCA TXOP. It is not granted in case that the TXOPLimit has been reached. Therefore, a station can transmit several frames for a particular AC without performing a backoff or waiting longer than SIFS before a transmission, realizing short packet bursts.

AC	CW_{min}	CW_{max}	AIFSN	AIFS [sec]	TXOPLimit [msec]	UP
0	31	1023	2	5.0E-05	0	1,2,0
1	31	1023	1	3.0E-05	3.008	3
2	15	31	1	3.0E-05	6.016	4,5
3	7	15	1	3.0E-05	3.008	6,7

Table 5.3: *EDCA Parameter Set*

The EDCA parameter set is shown in Table 5.3 [IEE03d]. An alternative parameter set is summarized in Table 5.4 [IEE03e]. This alternative set is used in the remainder of this work, since the interframe spaces of the EDCA parameters of Table 5.3 for ACs 1, 2, and 3 equal PIFS. Since PIFS is used before the transmission of Beacons, and Beacon frames are used by the AP to distribute management information, no IFS for data frames should be less than or equal to PIFS. A PIFS equal to AIFS increases the probability of a Beacon collision with data frames.

AC	CW_{min}	CW_{max}	AIFSN	AIFS [sec]	TXOPLimit [msec]	UP
0	31	1023	7	1.5E-05	0	1,2,0
1	31	1023	3	7.0E-05	0	3
2	15	31	2	5.0E-05	6.016	4,5
3	7	15	2	5.0E-05	3.264	6,7

Table 5.4: *Alternative EDCA Parameter Set*

HCF Controlled Channel Access (HCCA)

The HCF Controlled Channel Access mechanism is a polling mechanism similar to PCF. It defines a centralized coordinator, called Hybrid Coordinator (HC), which operates under QoS-aware rules with some significant differences to the Point Coordinator mechanism of PCF. The HC is collocated with the QoS enhanced AP and uses the PC's higher priority to gain access to the wireless medium to initiate frame exchange sequences for itself or to allocate TXOPs for QSTAs. The higher priority access to the medium is simply achieved by waiting for PIFS before accessing the wireless medium.

The HCCA mechanisms operate under both CFP and CP, in contrast to the PCF, which only operates under CFP. During CP, HCCA can allocate polled TXOPs for a QSTA to provide a limited-duration Controlled Access Phase (CAP) to transfer QoS data as shown in Figure 5.13. QSTAs use the virtual carrier sensing mechanism NAV during a CAP to protect ongoing HCCA transmissions.

HCCA operates during both CP and CFP to meet QoS requirements for different ACs. The HC medium access is defined by a QoS policy which can be specific for a particular BSS. Another significant difference to the PC is that the HC grants a QSTA a polled TXOP with a duration specified in the poll frame enabling the polled station to transmit

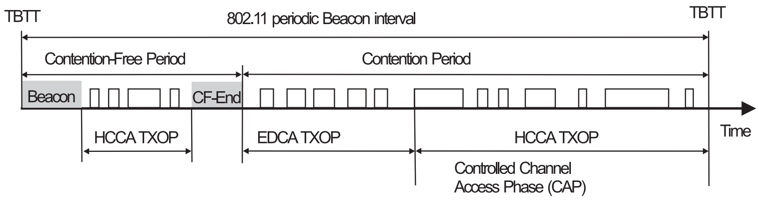

Figure 5.13: *HCCA and EDCA Transmission Opportunities*

multiple frames within the given polled TXOP. Like in PCF, the HC may perform a backoff before transmitting a poll frame under certain circumstances which are beyond the scope of this work.

The problems of this polled HCCA mechanism can easily be seen. Like in DCF, stations associate with the BSS during the CP. Therefore, if many CAPs are scheduled in a CP, no station can associate. This is similar to the problem with PCF, if the CP is assigned very small periods of time of the Contention-Free Repetition Interval.

Another problem is that CAPs could be deferred, if e.g. a non-QSTA is transmitting data when a CAP is to be issued. Deferred CAPs can lead to the problem that a CAP is not issued because the CP is finished. Thus, the HC cannot guarantee that a particular QSTA gets its CAP which it applied for.

Finally, HCCA has the same deferred Beacon problem as PCF. All these problems evolve for single cell scenarios. The deficits of polling mechanisms in more complex cell scenarios like overlapping or co-located cells have already been discussed in the last section. These results can be generalized for basically all MAC layer polling protocols as QoS enabler for wireless environments.

5.3 MAC Protocol Performance Evaluation

The last sections presented a detailed introduction to the different MAC protocols of Wireless LAN. The goal of our studies is to evaluate the QoS capabilities of the various mechanisms in a 4G environment. Such an environment necessitates the support of different QoS levels for the involved WLAN stations.

A detailed simulation was implemented using the OPNET® simulator. This includes the different MAC protocols, DCF, PCF, and HCF, as well as the applications as described in Chapter 3. Today, the IEEE 802.11b standard is the most widely used. Therefore, our simulations accounted for its physical layer with data rates of 1, 2, 5.5, and 11 Mbps in the 2.4 GHz frequency band using DSSS modulation.

This section is structured as follows. In Section 5.3.1 performance results of the basic IEEE 802.11 DCF MAC protocol within single cell scenarios are shown. This includes simulation studies about the effect of the number of Web users on the performance of Wireless LAN. These results indicate that the overhead of the RTS/CTS mechanism in cases with Hidden Node is too large to increase the system performance. Therefore a detailed study of the tradeoff between overhead and performance increase is given.

Then, Section 5.3.2 and Section 5.3.3 present the performance results for the more complex scenarios. The terms overlapping and co-located cells are introduced and the simulation scenarios with their involved applications are discussed. All of these cases are used to study the 4G environments and their impact on Wireless LAN performance.

5.3.1 Single cell scenarios

To evaluate the performance of a Wireless LAN, simple simulation scenarios with only one access point are considered and all the involved stations perform their transmissions using the basic DCF operating mode.

The results are used as a reference for the results of the more complicated scenarios later in this chapter. The basic scenario is shown in Figure 5.14. It has a single access point surrounded by a number of stations.

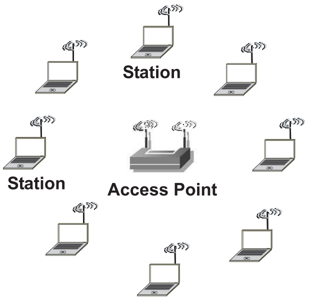

Figure 5.14: *Basic single cell scenario*

The number of stations is varied in the following scenarios. In addition, Hidden Nodes are introduced in two different ways. First, the stations form two different groups of stations, where the stations of the two groups are hidden from one another, while the stations within each group still receive each others' transmissions as indicated in Figure 5.15. The second way is to define all the stations as hidden from one another. In this case, a station can only receive the signal of the access point, but not of any other station.

The reason for the two cases is, that the two groups of stations present a rather realistic scenario. In practical environments, there is often the case where due to e.g. the structure of a building, there are clients that form subnet "islands". However, in order to study the effect of Hidden Nodes, the second case is better, since the effect of the Hidden Nodes becomes more obvious as the number of involved hidden stations is increased.

Basic performance evaluation of IEEE 802.11b DCF

Wireless LAN operating in DCF mode does not support any kind of service differentiation. All involved stations equally have to contend for the medium as is the case of Ethernet-like networks. The medium access is fairly shared between all stations. In such a scenario, the most widely used application is certainly the World Wide Web, as explained in Chapter 3. A Web user does not completely utilize the system bandwidth, but merely retrieves data whenever a new web page is loaded.

Therefore, the utilized bandwidth is no good measure of the performance of the system. A much better way to estimate the user-experienced quality is the average page download time. The higher the utilization of the system, the longer the page download times gets until a certain level is reached, where most users are dissatisfied. According to the Web source traffic model as described in Section 3.1.2, the average page size greatly varies. Therefore, the important measure is the relative behavior of the average page download time rather than the absolute value. A large number of users in this kind of simulation results in an increase in the delay experienced by each user. Very long delays cause TCP timeouts to expire and, therefore, lead to TCP retransmissions once a certain delay limit is reached. Any further increase of the cell load causes TCP to exceed its maximum number of retransmissions, such that the TCP connection is dropped and the page download is canceled ([Ste94]). Such dropped downloads do not contribute to our statistics. Therefore, the number of possible TCP retransmissions is set to unlimited.

Various performance studies of the Wireless LAN MAC protocol can be found in the literature, as in [BCG02], [KEW00], [RAHE01], [VBG00], or [AMC^{+}99]. These publications, however, focus on the properties of the MAC protocol itself, such as the maximum achievable throughput or fairness, but they ignore application-specific influences, which are of great importance to the subjective quality experienced by single users, e.g. in differing cell-load situations.

In the following, we study the effect of the number of concurrently active Web users on the system performance and the user-experienced quality of service ([Hec03a]). This does not only include the question of how many Web users can be served with adequate quality within a single cell, but it also shows the effect of mechanisms that were defined to overcome problems found solely in the wireless environment, such as the Hidden Node problem, and how these extensions affect the system performance.

The goal of these studies is to provide Wireless Internet Service Providers (WISP) with a better understanding of the capability of their WLAN infrastructure. We draw conclusions about the realistic capacity of single WLAN cells, which allow a better planning of WISPs' Internet access networks.

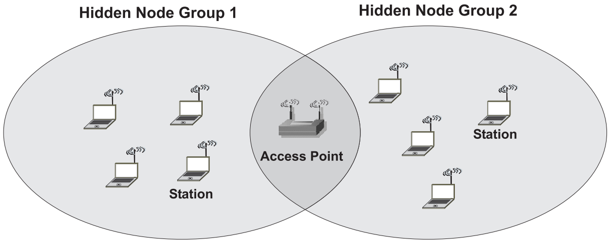

Figure 5.15: *Multiple Hidden Node groups*

The basic simulation scenario has already been shown in Figure 5.14. Here, we assume that all stations are in reception range of each other. In the following this case is referred to as the *single Hidden Node group*. In addition, simulation scenarios as shown in Figure 5.15 are considered in order to evaluate the effect of Hidden Nodes on the performance results. They consist of two groups of nodes that are hidden from one another. This case is referred to as *two Hidden Node groups*.

In the following, the RTS threshold is set to 256 Bytes, which means that if packets larger than the threshold have to be send, an RTS packet is issued prior to the data transmission. The WLAN standard allows fragmentation thresholds in the range of 256 Bytes up to 2312 Bytes. Only packets larger than the threshold are fragmented. In our simulations a fragmentation threshold of 256 Bytes is used if fragmentation is explicitly considered. In any other case, the fragmentation threshold is set to 2312 Bytes, meaning that it is turned off.

Figure 5.16 shows the results for the 1 Mbps scenario. The two solid lines represent the results for pure CSMA/CA and a single group of clients (no Hidden Nodes) and the case with two groups of users (with Hidden Nodes). The average page download time increases from approximately 0.6 seconds for the 10 clients to more than 10 seconds for 100 clients. This increase corresponds to a factor of more than 16, which means that a user in the 100 client case experiences page download times 16 times longer than for 10 clients. Such an increase is not acceptable, which means that the maximum number of Web users in the 1 Mbps scenario should not exceed 40 clients. Comparing the two curves yields the degradation of the system performance due to the Hidden Nodes. The gray line for the Hidden Node case is only about three percent above the black curve. Thus, the Hidden Nodes have a small but noticeable effect.

The dashed-dotted lines in Figure 5.16 show the results for the case that the RTS/CTS mechanism is activated. The average page download time for this case is always found to be about 10 percent above the scenario without RTS/CTS. This is true for the one and two groups scenarios. As we have discussed earlier, the RTS/CTS mechanism lowers the number of collisions in the 2 groups scenario (gray curves). However, it produces more overhead in our cases than can be gained by decreasing the probability of collisions.

Finally, the dashed lines correspond to the cases with additional fragmentation of packets larger than 256 Bytes. In our case, the wireless

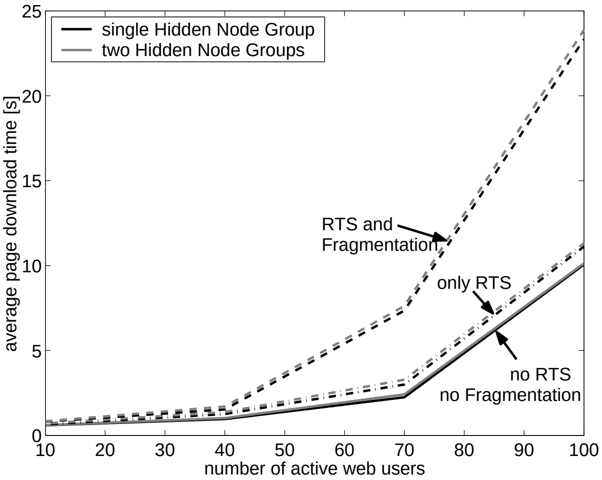

Figure 5.16: *Wireless LAN performance at 1 Mbps*

channel was assumed to be free of errors. The results, therefore, display the overhead introduced by fragmentation. It can be easily seen that the page download times are by far greater and that the fragmentation overhead has a major effect on the overall system performance.

The case where the maximum data rate is set to 11 Mbps as displayed in Figure 5.17 yields similar results. The Hidden Nodes (gray lines) cause a performance degradation of no more than three percent compared to the case without Hidden Nodes (black lines). The RTS/CTS mechanism overhead reaches about 10 percent and does not improve the overall performance, but leads to a further increase of the average page download times. The situation changes drastically, once the fragmentation mechanism is activated. The page download times almost explode and the WLAN cell can not handle more than 40 clients appropriately.

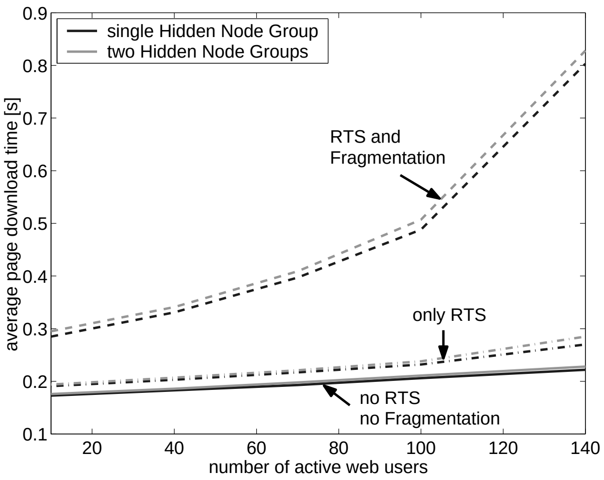

Figure 5.17: *Wireless LAN performance at 11 Mbps*

Nevertheless, the results in Figure 5.17 show that in the 11 Mbps case the system can easily handle up to 140 clients as long as fragmentation is not used. The average page download time for the 140 client case is less than 20 percent above the 10 client case. Considering the fact, that most currently available access points cannot support 140 simultaneously attached clients, we can conclude that in practice the performance of the system is still good enough to satisfy the Web users' demands even in environments with high cell loads.

Hidden Nodes and RTS/CTS

As explained earlier, the RTS/CTS mechanism was introduced to the Wireless LAN standard in order to solve the Hidden Node problem. It decreases the collision probability, such that a performance gain in terms of user experienced throughput can be reached. On the other hand, it adds to the overhead of the MAC protocol. However, it is expected that the gain is larger than the overhead, and thus an overall performance enhancement is reached. However, the results from the last section indicate that the overhead is too large, such that the performance is even more decreased when using the RTS/CTS mechanism in Hidden Node scenarios. These results are also approved by earlier studies as in [BCG02], [XS02], [Hec03a], and [Hec03b]. Therefore, we focus on this topic and discuss the various parameters that influence the tradeoff between additional overhead and performance enhancement in this section (see [HPW04]).

Again, the simulation scenario shown in Figure 5.14 is chosen. A number of WLAN stations surround the access point. Two different cases are considered. In the first case, all the involved WLAN clients are in the reception range of each other, such that no Hidden Nodes are present. Comparing the simulation results for the cases with and without the RTS/CTS mechanism yields the pure overhead caused by the MAC protocol extension.

In the second case, all the Wireless LAN stations are considered to be hidden from one another. This is accomplished by ignoring all but the access point's signal at the receiving clients. Again, the Wireless LAN performance can be studied with and without the RTS/CTS mechanism.

Several MAC protocol parameters can be adapted in order to increase the potential gain caused by RTS/CTS. One such parameter is the packet size. The larger the packet size, the larger the probability of a collision in the case without RTS/CTS. Therefore, the packet sizes 2312 Bytes and 1500 Bytes are used in our simulations. The simulation scenario neglects all effects caused by a wired backbone, which is the desired behavior. The data rate is set to 11 Mbps and we assume that the signals can be received without any bit errors in the close distances assumed in the scenario.

A number of different applications is assumed. For the first set of simulations, it was assumed that all stations are saturated sources of UDP traffic, i.e. each WLAN station has always data to transmit using the maximum packet size of 2312 Bytes. Each single client can, thus, use up all the available system bandwidth. The access point does not transmit any data to the stations, such that only uplink traffic is assumed. The results are shown in Figure 5.18. It depicts the average overall throughput with 90 percent confidence intervals. The abscissa shows the number of involved clients, and the ordinate represents the throughput in Mbps.

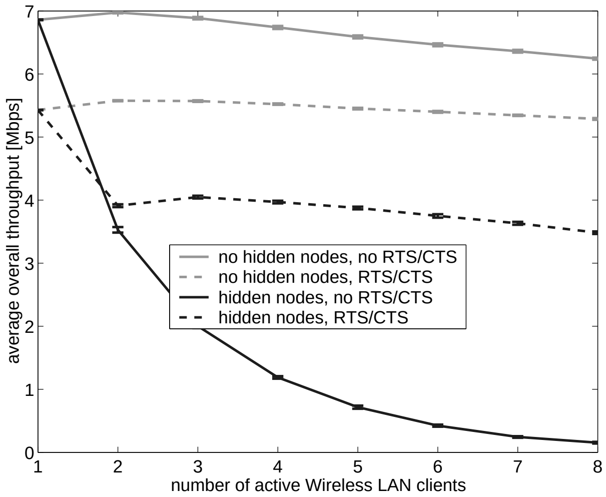

Figure 5.18: *Average UDP uplink throughput at the access point using maximum packet size*

The two gray lines show the case where all the involved clients are within the reception range of each other, i.e. no Hidden Nodes are present. The solid gray line is the case where the RTS/CTS mechanism is deactivated. We can see that a maximum throughput of about 7 Mbps can be reached in this case. Comparing it with the dashed gray line, where RTS/CTS is turned on, yields the overhead of RTS/CTS. In this case we can see that the overhead is approximately 20 percent.

The black lines, on the other hand, show the case where all the involved clients are hidden from one another. The solid black line is the case with no RTS/CTS, while the dashed black line has RTS/CTS turned on. It can be seen easily that even for just two Hidden Nodes, the RTS/CTS mechanism improves the system performance. As the number of Hidden Nodes increases, the dashed black line decreases only slightly, but still a good performance can be reached (around 4 Mbps average throughput). If RTS/CTS is not used, the performance drops rather quickly to almost zero. Here, we can clearly see an advantage of RTS/CTS in Hidden Node environments.

To study the effect of the packet size the same simulations of UDP upstream traffic were performed with a packet size of 1500 Bytes. This is a much more realistic scenario, since the maximum allowed packet size of 2312 Bytes is hardly ever seen in a normal environment, where Wireless LAN is combined with Ethernet. The results are shown in Figure 5.19.

First of all, it can be derived that the maximum performance drops from almost 7 Mbps to just about 6 Mbps. This 15 percent performance degradation is solely based on the smaller packet sizes and the increased overhead. Then, we can see that the overhead of the RTS/CTS mechanism in the case with no Hidden Nodes, as shown by the two gray curves, stays about the same. It is again about 20 percent. The black curves, which again represent the case where all the clients are hidden from one another, show a similar behavior to the case with maximum packet sizes. However, this time the dashed black line crosses the solid black line for values greater than two. This means, that if only two Hidden Nodes are present, the case without RTS/CTS still outperforms the case with

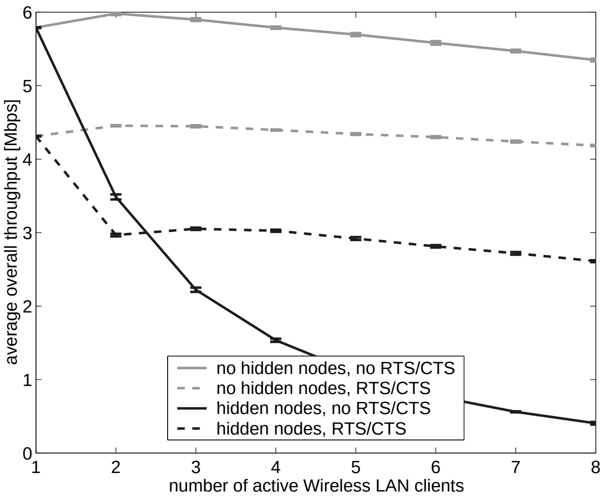

Figure 5.19: *Average UDP uplink throughput at the access point using 1500 Bytes packet size*

RTS/CTS. For more than two clients, the dashed black line again drops gradually, while the solid black line shows a fast decrease to almost zero. The system performance in the case of RTS/CTS is, thus, less affected by further increasing the number of Hidden Nodes.

UDP traffic on the downlink was not considered, since it is not at all influenced by the number of Hidden Nodes in the system. Traffic only gets transmitted from the access point to the different clients. Collisions never occur. Therefore, the maximum throughput in the case of UDP downlink traffic is decreased by about 20 percent when utilizing RTS/CTS, since no performance gain can be achieved and only additional overhead is induced.

Now, FTP traffic is considered. We distinguish the cases with only uplink traffic, only downlink traffic, and a mixture of uplink and downlink traffic. Stations do not have any idle periods, i.e. as soon as a download or upload is finished, the next one is started immediately. It should be pointed out that a fairness problem occurs in the case of FTP traffic. It turns out that in all Hidden Node cases where pure FTP uploads or the mixture of FTP up- and downloads are considered, all but one FTP (TCP) connection setup fails and therefore only one of the Hidden Nodes gets served. However, such a behavior is not acceptable. Using the RTS/CTS mechanism does not at all change the situation. The problem is that the probability of a collision is large. The number of retries in setting up a new FTP/TCP connection is rather small. Therefore, the probability of a collision has to be further diminished by changing the Contention Window size. In the case of pure FTP upload and mixed FTP up- and downloads, the CW parameter CW_{min} is set to 255 instead of 31, while the CW_{max} value is left untouched at 1023.

In all of these cases, the measure of choice to evaluate the performance of the system is throughput in Mbps. It is calculated as the sum of all the data packets correctly received by any station or by the access point supposing the packet was addressed to the station or access point. The RTS/CTS packets are not considered. Therefore, our measure yields the overall system performance.

When considering pure FTP downloads, merely downlink traffic from the access point to the Wireless LAN stations occurs. Therefore, similar results to the case with pure UDP downlink traffic can be found as shown in Figure 5.20. We can see that there is two pairs of coinciding curves. The first pair consists of the two dashed lines, where RTS/CTS is turned on. The other pair, consisting of the two solid lines, has RTS/CTS turned off. The number of Hidden Nodes does not have any impact on the system performance. The RTS/CTS mechanism just adds to the overhead of the MAC protocol. The performance drops by about 20 percent.

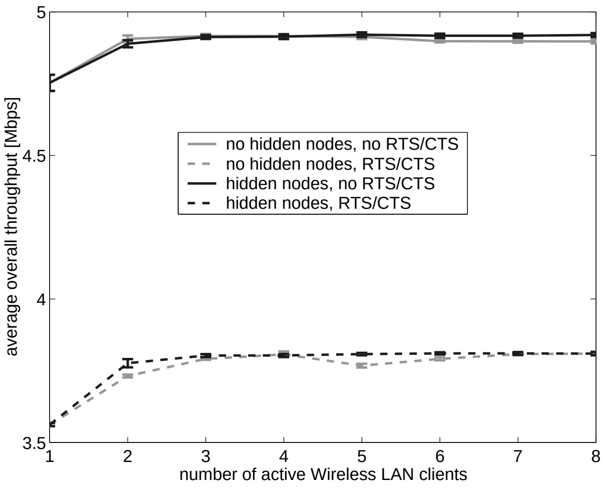

Figure 5.20: *Average throughput in the case of FTP downloads*

The situation changes completely once we consider FTP upload traffic. First of all, the Contention Window parameters have to be adapted to the situation as explained above. Due to the increase of the collision probability, all but one Hidden Node are not able to successfully set up a TCP connection. Consequently, only one station can exclusively commit an upload, which is definitely not an acceptable system behavior.

After adapting the Contention Window parameters, all the stations are successful in uploading files to the FTP server, virtually located in the access point. However, in this case the Hidden Nodes have a remarkable influence on the average throughput of the system as shown in Figure 5.21. In the case where no RTS/CTS is used, the performance drops from just above 3.4 Mbps to less than 1.8 Mbps, which is a decrease of almost 50 percent. On the contrary, if RTS/CTS is turned on, the throughput is decreased from around 2.8 Mbps to just above 2.4 Mbps, which is a decline of less than 15 percent.

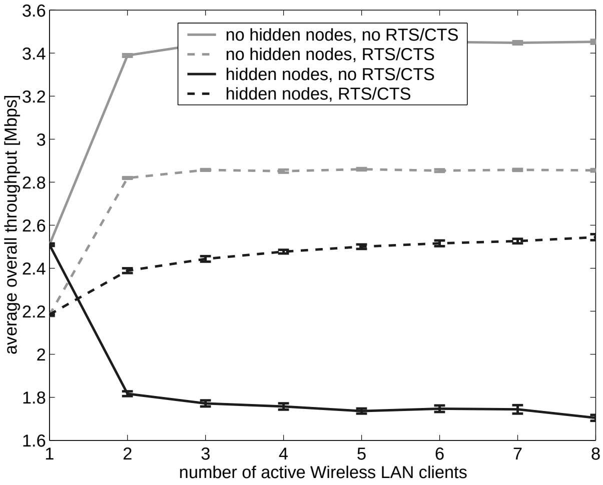

Figure 5.21: *Average throughput in the case of FTP uploads*

Another interesting fact can be seen in Figure 5.21. The achievable throughput increases with the number of active stations in most cases. This is caused by the backoff algorithm that is simultaneously performed by all stations, i.e. the backoff slots are simultaneously decreased in all stations. The average packet interarrival time as seen by the access point is, thus, smaller than the mean number of backoff slots of each client.

Nevertheless, it is important to notice that a scenario with only FTP upload seems rather hard to find in reality. Usually, there is much more downlink traffic than uplink traffic. Therefore, a more realistic case of an equally shared mixture of FTP uplink and downlink traffic is considered next. The results are shown in Figure 5.22.

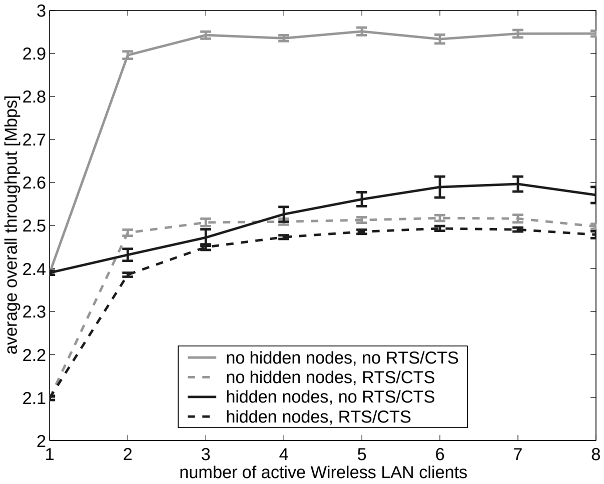

Figure 5.22: *Average throughput with mixed FTP uplink and downlink traffic*

Most of the curves exhibit a similar behavior as the results for pure FTP uplink traffic. The average throughput increases with the number of active stations (gray solid line and dashed lines). However, in this case, the performance for the Hidden Nodes case without RTS/CTS, the solid black line, shows a completely different shape. While it dropped dramatically for an increasing number of Hidden Nodes in the previous scenario, it increases gradually in this case. It always lies above the dashed black curve, which means that even in the case with Hidden Nodes, the performance of the system is better when RTS/CTS is turned off. Interestingly, this is also true comparing the solid black line and the dashed gray curve. Even in the case where there are eight Hidden Nodes, the system without RTS/CTS shows a better performance than in any RTS/CTS case.

Summing up the FTP scenarios, we can conclude that there is only one case where RTS/CTS can improve the system performance in terms of throughput. That is where solely FTP upload traffic is simulated. Considering the low practical relevance of such a scenario compared to the cases of pure FTP downloads and a mixture of uplink and downlink traffic, we are clearly inclined to prefer a system configuration without RTS/CTS. On the other hand, we have to keep in mind that there is an unfairness problem in the FTP case which made it necessary to adapt the Contention Window parameters.

Finally, Web users and HTTP traffic are studied. It is the most realistic scenario considered in our RTS/CTS studies. Web users are by far the most common traffic sources nowadays, especially in wireless environments. However, a Web user significantly differs from the UDP and FTP users described above. They do not consume all the bandwidth offered by the system. Therefore, a different kind of measure has to be chosen in order to evaluate the system performance in this case. So far, we considered the overall average throughput reached by the system. This was a good choice, since all the clients had enough traffic to use up the complete system bandwidth. In the case of Web users, the situation changes completely. Each user just has a demand for less than 10 Kbps on average. Therefore, system bandwidth is not used up completely even for large numbers of users, and the maximum throughput of the system cannot be retrieved here. A good measure for this type of simulation is the average page download time for the involved Web users. It clearly states the subjective quality of service that each user experiences. On the other hand, it allows to compare the results for the different test cases.

The results are shown in Figure 5.23. It can be seen that an increase in the number of users leads to a linear decline of the performance for the two cases where RTS/CTS is turned off, the two solid lines. Otherwise the two dashed lines with RTS/CTS switched on, indicate that the system

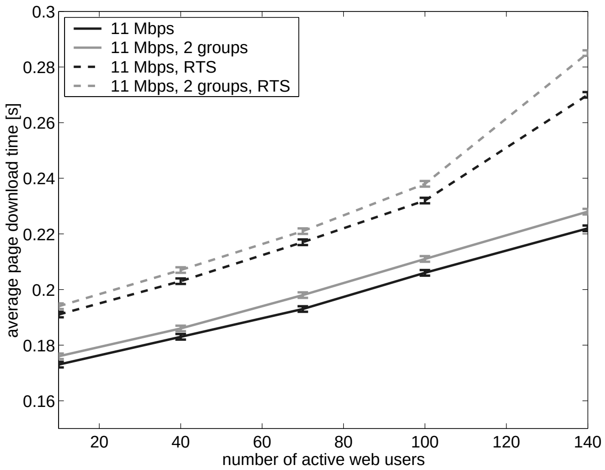

Figure 5.23: *Subjective page download time experienced by Web users*

almost reached its limits and a further increase in the number of users leads to an explosive increase in page download time.

Comparing the two solid lines, we can conclude that the Hidden Nodes lead to a performance decrease of less than three percent, which can hardly be noticed by the user. On the other hand, the RTS/CTS cases all exhibit a by far worse behavior. No matter if there are Hidden Nodes in the simulated environment or not, the performance is degraded by more than 15 percent.

Summing up all the results found in the RTS/CTS and Hidden Node study, we can conclude that the results greatly depend on the type of traffic that is simulated. In cases where the uplink traffic is predominant, RTS/CTS can clearly improve the system performance in Hidden Node scenarios. However, the more downlink traffic occurs, i.e. the higher the practical relevance of the studied scenario, the worse does the RTS/CTS mechanism perform.

Another important point is that it is hard to find practical Wireless LAN scenarios with up to eight Hidden Nodes. This is definitely an exceptional case. Our experience indicates that the number of Hidden Nodes in a single cell is hardly greater than two. This again adds to the conclusion to leave RTS/CTS turned off.

After a number of single cell scenarios have been analyzed in this section, we turn to more complex simulation scenarios involving multiple cells in the next section.

5.3.2 Overlapping and co-located cells in DCF mode

The last section presented results about Wireless LAN performance studies in single cell scenarios. Various other performance studies of the Wireless LAN MAC protocol can be found in the literature, e.g. [CGL00], [GZ03], [CWKS97], [LAS01b], and [LAS01a]. These publications, however, focus on MAC protocol performance issues within single cell scenarios. This is not sufficient for the evaluation of Wireless LAN as a future access technology in 4G networks. Therefore, we investigate the impact of *overlapping* and *co-located cells* on the performance of the different Wireless LAN MAC protocols in this section. We evaluate how the users in different cells interact and what the consequences are on the performance. We identify situations where the communication of single clients is completely blocked due to high collision probabilities and the unfairness in distributed environments.

Simulation Scenarios

In order to study the effect of multiple cell scenarios on the performance of the MAC protocol, we have to identify the scenarios where such effects might arise. Due to the small number of non-overlapping channels in

Wireless LAN, there might be situations where an overlap of cells in terms of coverage and channel can not be avoided.

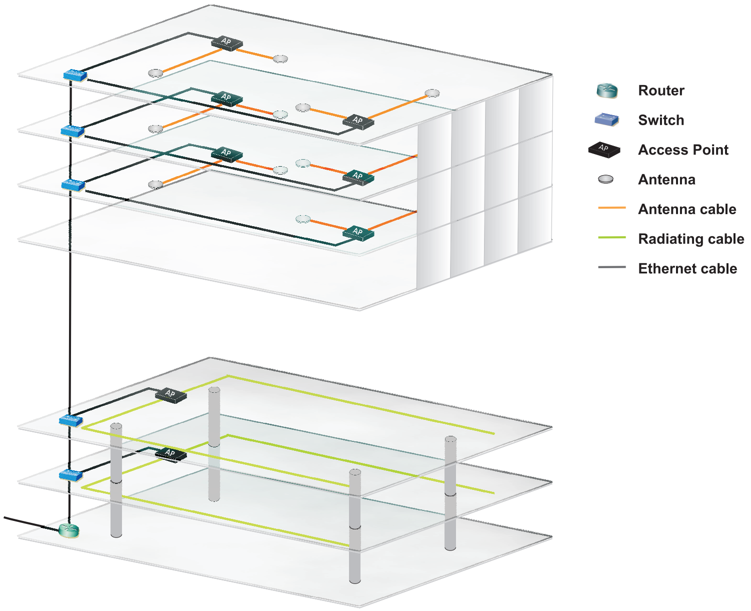

Figure 5.24: *Wireless LAN deployment in a large office building*

Consider the scenario shown in Figure 5.24. It shows a large office building that is completely covered by Wireless LAN. However, IEEE 802.11b only supports three non-overlapping channels, as we have already seen in Section 4. This means, that overlaps occur and the technology has to make sure that it can deal with these situations. This is especially necessary if QoS demanding applications, such as VoIP, are used.

The IEEE 802.11a standard which utilizes the 5 GHz band, supports up to eight non-overlapping cells. Nevertheless, the signal attenuation in the higher frequency is by far larger and therefore, the IEEE 802.11b standard is usually better suited for indoor usage, not considering special cases. If the IEEE 802.11a standard is to be used indoors, a much higher access point density has to be used, which again leads to a higher number of channels necessary to avoid overlaps in coverage and channel. Thus, even in these cases overlaps occur, and the MAC protocol has to take care of it.

In the following, we consider simulation scenarios that consist of two access points (A1 and A2) and a single Wireless LAN station within each of these cells. Station C1 is communicating with access point A1, while station C2 in connected to access point A2. There are seven different scenarios that can be defined in this case. We distinguish the cases where the two access points are not within reception range of one another, the overlapping cells, and the cases where the access points are in reception range of each other, the co-located cells.

The planning process of a Wireless Internet Service Provider tries to avoid co-located cells, since they do not exploit the maximum coverage that the two access points can reach. Placing the access points farther apart would usually lead to a larger covered area. However, there are situation where this is not possible, such that co-located cells definitely have to be dealt with by the MAC protocol.

In the case of *overlapping cells*, there are three possible scenarios as shown in Figure 5.25. The first overlapping cells scenario is marked with an A. It shows the coverage areas of the two access points A1 and A2 as gray solid circles around the nodes. The two Wireless LAN stations C1 and C2 are placed in the coverage area of both access points. In this scenario both clients experience the same problems caused by the overlap. The reception range of the two clients is indicated by the dashed gray circles.

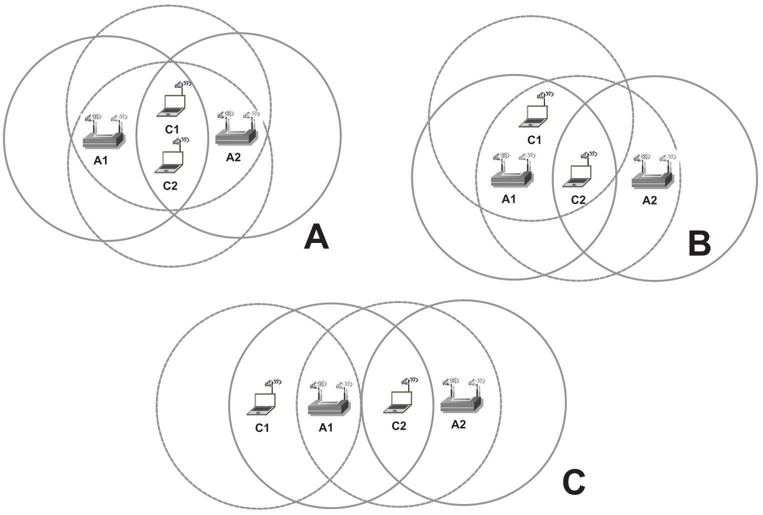

Figure 5.25: *Multiple cell simulation scenarios: overlapping cells*

Scenario B changes the position of client C1. It is not in the reception range of the access point A2, but still receives the packets transmitted by the other client C2. The client C2 is still in the coverage area of both access points. Finally, in scenario C the client C1 is placed farther away from the access point A2 and the client C2. It is now only in the reception range of its associated access point A1. The client C2 is still located in the area covered by both access points.

The different scenarios that can be found for *co-located cells* are shown in Figure 5.26. In all these cases, the access points are in the reception range of each other. An appropriate planning process should try to avoid these situations, but as more wireless operators start their service while private users set up their own private hot spots, these scenarios are

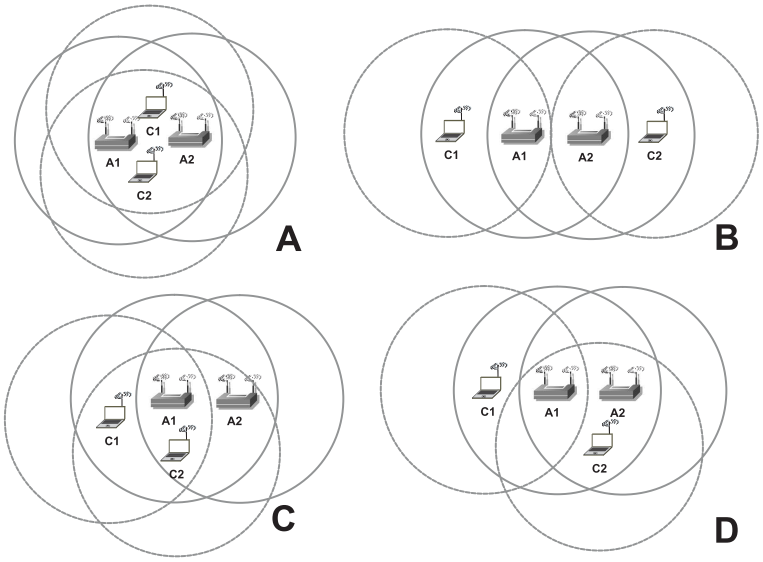

Figure 5.26: *Multiple cell simulation scenarios: co-located cells*

definitely possible in practice. The Wireless LAN MAC protocol should still be able to serve the users in a fair manner and with the assigned QoS level.

The scenario marked with an A shows the case where all the involved stations are placed in close vicinity. Each node receives the transmissions of all the others. This case should not be too different from the case where two clients are located in the vicinity of a single access point, in terms of the bandwidth they can receive. Scenario B, on the other hand, shows the case where the two clients C1 and C2 are located in the reception range of their own but not of the other access point. The clients' transmissions are not disturbed by the access points. However, the access points disturb each other's signals.

In scenario C, the station C1 is located outside the reception range of access point A2, while station C2 can receive the signals from all the involved nodes. Scenario D is a slight modification of scenario C. Here the client C2 is moved away from the client C1, such that client C2 is outside the reception range of client C1. However, it is still in the area covered by both access points.

In order to evaluate the effect of the overlapping an co-located cells, reference cases are needed. In our studies we define three different reference scenarios that allow us to study the influence of the location of the various stations and access points on the performance of the Wireless LAN cells. These scenarios are shown in Figure 5.27.

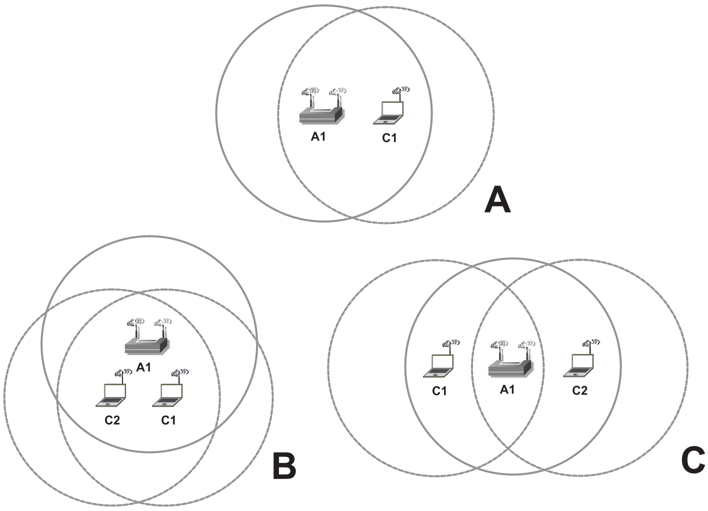

Figure 5.27: *Multiple cell simulation scenarios: reference scenarios*

Scenario A consists simply of one access point and a single station. It is used to derive the maximum performance of the Wireless LAN MAC protocol in the absence of multiple stations and access points. Scenario B involves a second station. All stations are located in the reception range of each other. This helps to evaluate the performance of a two station scenario without the influence of multiple cells. Finally, in scenario C the clients are located farther apart from each other. This helps to include the influence of Hidden Nodes, a common case in multiple cell scenarios.

Fairness Indicator

One of the main goals of a Medium Access Control protocol is to achieve fair access to the medium for all involved stations. If QoS aware applications are used, a prioritized access might be implemented, but stations operating on the same QoS level should be treated in the same way, i.e. fairness has to be guaranteed.

Nevertheless, there might be cases where the average throughput is shared equally over longer periods of time, such that fairness in terms of average throughput is given on large timescales. However, if we look closer we might find that average throughput is not the only important factor. The stations do not always alternate their access on the medium as we might expect, i.e. fairness on smaller timescales might not be given. The following figures show different ways the two stations $C1$ and $C2$ get access to the medium. The bars depict representative time periods of 10 seconds. A gray line from top to bottom shows a successful packet reception of one station while the black lines represent the packet reception at the other station. Alternatively, depending on the simulated scenario, the gray and black lines could also represent a successful packet reception at the stations' associated access point.

Figure 5.28 shows a fair sharing of the medium. Over the whole 10 second period the two stations alternatingly receive packets. Small black

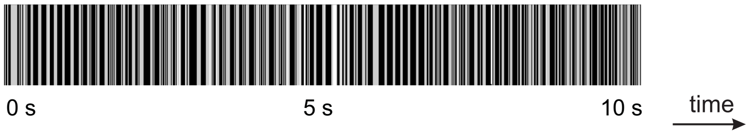

Figure 5.28: *Fairness Indicator: fair, alternating access to the medium*

or gray blocks indicate that the particular station received multiple packets. This result was found using reference scenario B. Conversely, Figure 5.29 shows the unfair counterpart. The alternating gray and black blocks show that the stations block each other over longer periods of time. For the first five seconds, one station exclusively utilizes the medium shown by the black blocks. Then, the situation changes and only the other client receives data packets for the remaining time shown by the gray blocks. Such an unfair behavior is not desired.

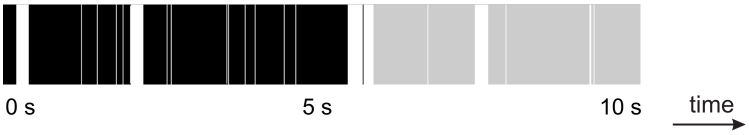

Figure 5.29: *Fairness Indicator: unfair sharing of the medium*

Therefore, in all our studies we have to account not only for fairness in terms of average throughput, but also in terms of alternating access to the medium.

Best-effort traffic performance

In this section, the contention-based Medium Access Control protocol as defined in the IEEE 802.11 and IEEE 802.11b standard is analyzed. The Distributed Control Function (DCF) is explained in details in Section 5.2.3. The main goal is to evaluate fair medium access in overlapping and co-located cell scenarios. The involved stations perform FTP downloads using different file sizes from the FTP server located on their associated access points as explained in Section 3.3. This simple case is used to better understand the impact of the multiple cell scenarios (see [Hec03b]).

To compare the results with the single cell cases, we first present the performance studies for the reference scenarios. The results are summarized in Table 5.5. It shows the average throughput in KBps (Kilobytes per second) experienced by the two Wireless LAN stations. The reference scenarios are all symmetric, such that both clients receive the same average throughput. Therefore, just one value is given for each case.

Scenario	**RTS**	10 KB	100 KB	1 MB	10 MB
A	-	80	520	714	751
	256	80	520	597	616
B		80	260	354	374
	256	80	260	312	303
C		79	258	347	358
	256	77	256	289	297

Table 5.5: *Reference scenarios: average throughput in KBps*

Table 5.5 shows that the throughput increases as the size of the requested file is increased from 10 KB to 10 MB. The results for reference scenario A show the maximum achievable throughput at about 750 KBps, because no other station is involved, such that the number of collisions is minimal. In addition, it shows the overhead induced by the RTS mechanism (an RTS threshold of 256 Bytes was chosen). It reaches a maximum of about 20 percent in the case of 10 MB file downloads.

The results for scenario B show the average throughput when two clients are simultaneously active in a single cell. In the case of small file sizes, the medium is under low load, such that both stations can be served like in the case with just one station. This is due to the dalay caused by the setup of the FTP connections. Each of the two stations receives about 80 KBps. As the file size, and thus the load increases, the clients still share the throughput equally. Under high load, the RTS mechanism causes an overhead of 20 percent.

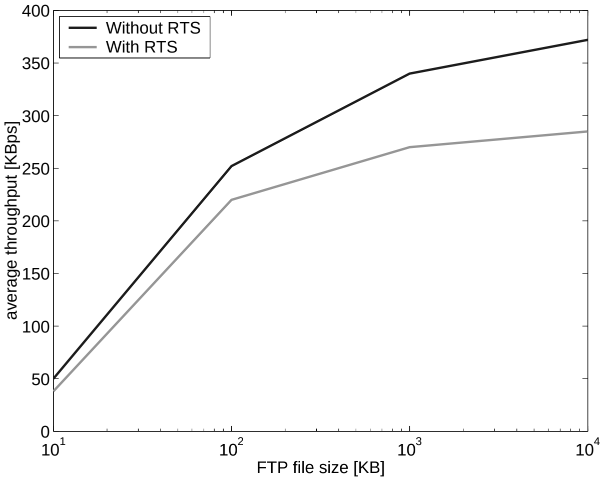

Figure 5.30: *Average throughput in the overlapping cells scenario A*

The last two rows of Table 5.5 show the impact of the two stations being hidden from one another. As long as the load is low, the average throughput can be kept at the same level, but as the load increases, the throughput declines by around five percent in the case without RTS and two percent with RTS.

In the reference scenarios, the Fairness Indicator shows that the two stations are served in a fair manner. They alternate their access to the medium. As an example Figure 5.28 is a representative time period for reference scenario B with a file size of 10 MB.

Let us now consider the simple case of the overlapping cells scenario A, where both stations are located in the overlap of the two access points. This is a symmetric case, such that the results are the same for both involved stations. Figure 5.30 shows the average throughput received by either of the two stations. By comparing the results to the reference scenarios, we can figure out that the achievable throughput is decreased by just a few percent, even though the number of collisions in this case is considerably higher.

However, the representative time period shown in Figure 5.31 indicates that the DCF MAC protocol does not guarantee fairness. While there are only short periods of time when both stations alternatingly receive data packets, there is a period of eight seconds where one of the stations almost exclusively utilizes the medium, which is not a desirable system behavior.

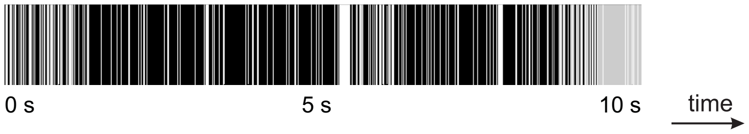

Figure 5.31: *Fairness in the overlapping cells scenario A*

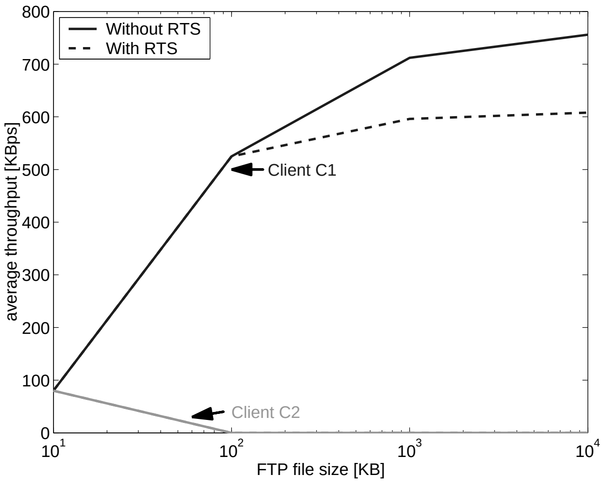

Figure 5.32: *Average throughput in the overlapping cells scenario B*

This unfairness is by far intensified once we consider overlapping cells scenario B. In this asymmetric case, only station C2 is in the overlap of the two access points, while station C1 is not disturbed by the data transmission of access point A2. Figure 5.32 shows that for an increasing load on the medium, the station C1 can take all the bandwidth it needs, while station C2 is only able to utilize the remaining bandwidth. For large file sizes this means that station C2 is not able to receive any more data. The figure also shows that using RTS/CTS does not at all ease the problem. The station C2 can still only utilize the remaining bandwidth.

As a solution to this problem, the chances of a successful transmission of station C2 have to be increased. This can be achieved by adjusting the Contention Window parameters appropriately. We introduce a set of different priority classes as shown in Table 5.6. The higher the priority class, the larger the Contention Windows, and thus the lower the probability of getting access to the medium.

Priority class	0	1	2	3	4	5	6
CW_{min}	7	15	31	63	127	255	511
CW_{max}	15	127	255	511	1023	2047	4095

Table 5.6: *Wireless LAN priority classes*

Alternatively, other MAC protocol parameters could be used to prioritize certain stations over others, such as the Interframe Spaces. However, it turned out that varying the Interframe Space has just little impact on the fairness in multiple cell scenarios.

Using the different priority classes in the overlapping cells scenario B yield the results shown in Table 5.7. Again FTP downloads from their associated access point were performed with the 1 MB file size.

The results show that for the cases without RTS/CTS, the only acceptable way to achieve a solution to the problem is to use the lowest priority class 6 at station C1 and the highest priority class 1 at station C2. The bandwidth is not shared equally. The station C1 still receives 264 KBps or about 63 percent of the available bandwidth, while station C2 can only use up 157 KBps or 37 percent of the available bandwidth. The drawback of this way of prioritization becomes obvious as well, because the overall bandwidth received by the two stations sums up to be only 421 KBps, compared to the 714 KBps maximum throughput received in reference scenario 1. This is a 40 percent reduction of the available bandwidth. However, it allows a fair sharing of the medium, which is the more important factor, since fairness is of highest priority to all MAC protocols.

On the other hand, Table 5.7 shows that when RTS/CTS is used, a number of different priority settings become possible. The priority classes

Priority Class C1	**Priority Class C2**	**RTS/CTS**	**throughput C1**	**throughput C2**
4	1	-	556 KBps	0 KBps
5	1	-	417 KBps	2 KBps
5	2	-	417 KBps	2 KBps
6	**1**	**-**	**264 KBps**	**157 KBps**
4	2	256	459 KBps	2 KBps
5	2	256	364 KBps	13 KBps
5	3	256	233 KBps	10 KBps
6	**3**	**256**	**204 KBps**	**163 KBps**
6	**4**	**256**	**224 KBps**	**108 KBps**
6	**5**	**256**	**237 KBps**	**70 KBps**

Table 5.7: *Overlapping cells scenario B with prioritization (1 MB files)*

(6,3), (6,4), and (6,5) lead to acceptable results. This means that the robustness of the DCF MAC protocol is by far increased if the RTS/CTS mechanism is used in multiple cell scenarios. In addition, such a prioritization leaves some higher priorities unused, such that there is still potential for higher priority classes, for example in the case of high priority traffic. Therefore, we set our focus to these parameters in the following.

Figure 5.33 presents the results for the different priority classes that can be used when RTS/CTS is applied. It can be seen that there is the option to use different priority settings to perform a fine-grained prioritization. While priorities (6,5) privilege station C1 in terms of throughput, a setting of (6,3) leads to more equal shares regarding the throughput rates.

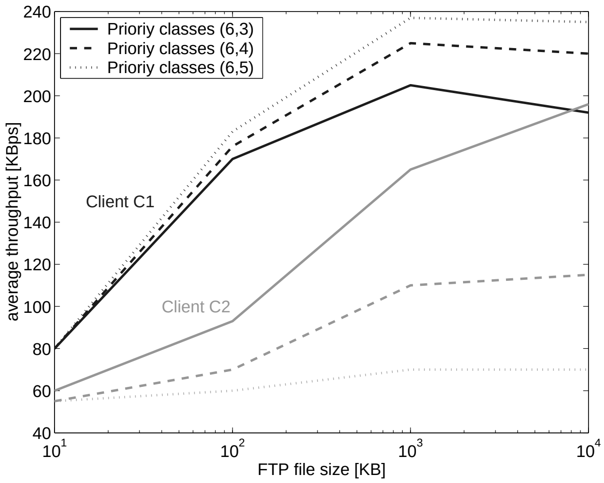

Figure 5.33: *Average throughput in the overlapping cells scenario B with prioritization*

The Fairness Indicator in Figure 5.34 shows that fairness is not just given on the average throughput basis, but also in terms of alternating access to the medium. The overall performance in terms of average throughput sums up to 367 KBps for the (6,3) priority classes, 332 KBps for the (6,4) case, and 307 KBps for (6,5) priority classes. This means that the maximum performance drops down to 51, 46, and 43 percent respectively, but considering the dramatic problems of the pure DCF mechanism in overlapping cells, this seems rather acceptable.

Applying these priority settings to the overlapping cells C leads to the results shown in Figure 5.35. In this case, we can conclude that a station in the overlap of two cells should increase its own priority to either class 4 or 5. In addition, it has to inform its associated access point to change

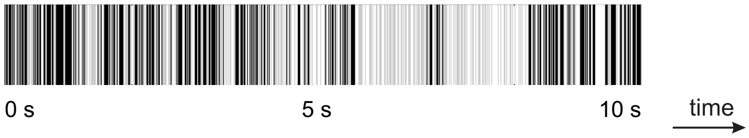

Figure 5.34: *Fairness in the overlapping cells scenario B with prioritization*

its Contention Windows settings. Stations that can only receive a single access point, meaning that they are not in an overlapping area, should use priority class 6.

The Fairness Indicator in Figure 5.36 shows that also in the case of overlapping cells C, the prioritization of the disadvantaged station leads to fair medium share. Most of the time, the two stations alternate their access to the medium as is desired.

Summarizing the results found for the overlapping cells, we can conclude that there are dramatic fairness problems if the standard DCF access mechanism is used. Stations in the overlap cannot receive an equal share of the medium compared to stations that are located outside of the overlap. In the worst case, i.e. if a station outside the overlap utilizes the whole bandwidth, a station that is located within the overlap but associated to the other access point can not access the medium at all. Therefore, these disadvantaged stations have to be prioritized in some way. It turned out that the Contention Window parameters CW_{min} and CW_{max} provide an easy and flexible way to implement such a prioritization. Applying different priority classes to the station depending on their location eases the problems. If the RTS/CTS mechanism is turned on, the robustness of the MAC protocol is further increased, and a number of different priority adaptations become possible.

Let us now turn to the co-located cells scenarios. The two symmet-

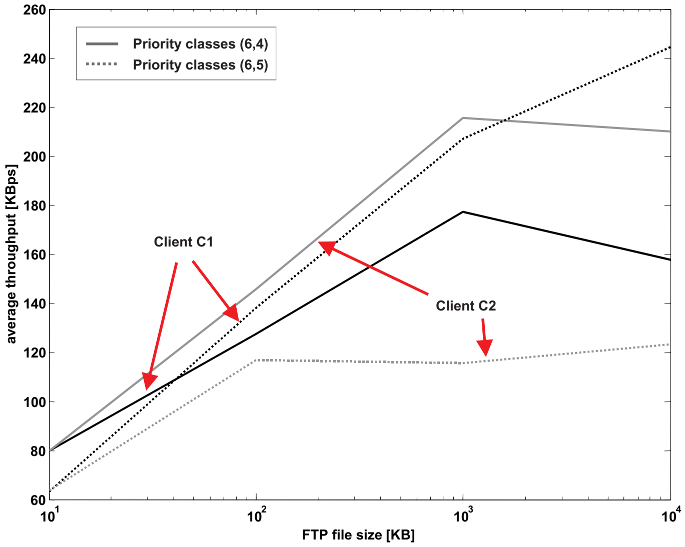

Figure 5.35: *Average throughput in the overlapping cells scenario C with prioritization*

ric co-located cells scenarios A and B do not cause any fairness problems. The simulations with the default Contention Window setting yield throughput rates that are comparable to the reference scenarios as shown in Table 5.8. Again, the maximum achievable throughput increases with the file size. In the case of 10 KB file downloads each station receives the full 80 KBps as in the reference scenarios. Once the file size is increased to 10 MB the bandwidth of the medium is completely utilized. However, in the case of co-located cells B, the maximum throughput for the 10 MB file size is 378 KBps for each of the two stations, compared to 428 KBps in the co-located cells A. The reason is that the two stations in the

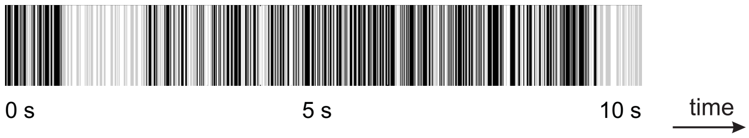

Figure 5.36: *Fairness in the overlapping cells scenario C with prioritization*

Scenario	**RTS**	10 KB	100 KB	1 MB	10 MB
A	-	80	263	418	428
	256	51	256	312	319
B	-	78	260	325	378
	256	51	256	312	319

Table 5.8: *Results for co-located cells A and B*

co-located cells B are hidden from one another, such that the number of collisions increases, which leads to a reduction of the bandwidth.

Nevertheless, the two symmetric co-located cells A and B do not cause any fairness problem. No corrective actions have to be taken.

However, for the co-located cells C, the situation changes dramatically. Figure 5.37 shows that one of the stations is experiencing an extreme unfairness. It can not receive any more data packets when the load of the cell reaches a certain level. The RTS/CTS mechanism does not ease the problem. In contrast to the overlapping cells, here the station C1 is disadvantaged. This means that in the co-located cells not the station within the overlap but the station in the reception range of a single access point has to be privileged.

Applying the priority classes as found for the overlapping cells but

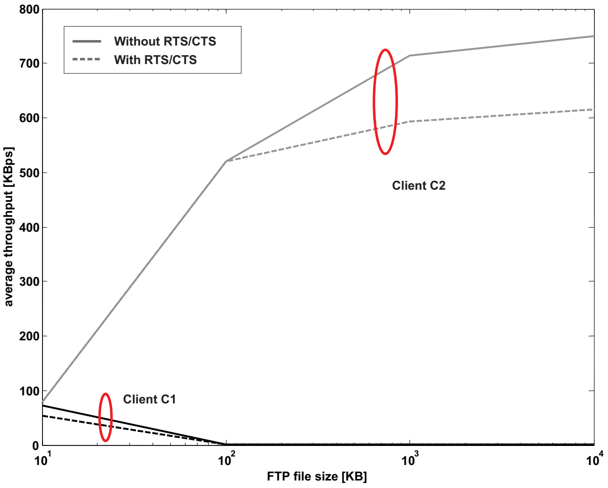

Figure 5.37: *Average throughput in the co-located cells scenario C*

in reverse order leads to the results shown in Figure 5.38. Again, the problem is solved and both stations share the throughput adequately. Also fairness is given as shown in Figure 5.39.

Similar results can be found for the co-located cells D as summarized in Table 5.9. Again, different priority settings can be used to perform a fine-grained prioritization if the RTS/CTS mechanism is used.

Combining the results for both, the overlapping cells and the co-located cells, the following mechanism can be proposed. For increased robustness of the DCF MAC protocol in multiple cell scenarios, the RTS/CTS mechanism should be turned on. The stations inform their associated access points whether they are in the coverage area of one or more access points. The Beacon frames that are transmitted by the access points on a regular basis, can be used for the location assessment

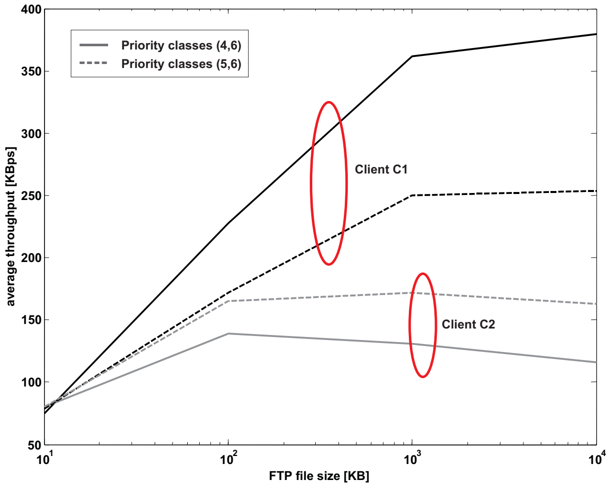

Figure 5.38: *Average throughput in the co-located cells scenario C with prioritization*

of stations. The access points, on the other hand, scan their channel for other access points in their reception range. Again this can be implemented by simply listening to the Beacon frames. If the access point finds itself in an overlapping cell, it tells all its associated stations that are in an overlap to increase their priority level to class 6. In the case the access point is placed in a co-located cell this process is inverted. Clients in the overlap are told to use the low priority class 4, while all others should use priority class 6.

There are situations where such a prioritization is not necessary, but the maximum performance can be reached if the default Contention Window settings are used. However, in order to provide a robust system and to keep the configuration tasks at a minimum level, this simple algorithm

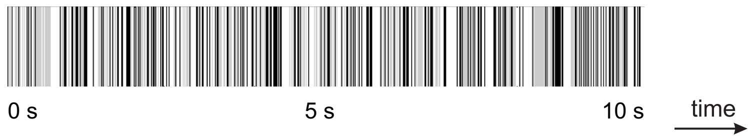

Figure 5.39: *Fairness in the co-located cells scenario C with prioritization*

Priority Class C1	**Priority Class C2**	**RTS/CTS**	**throughput C1**	**throughput C2**
5	6	256	297 KBps	134 KBps
4	6	256	415 KBps	78 KBps

Table 5.9: *Results for co-located cells scenario C (file size: 1 MB)*

should be applied. Implementing this solution improves the fairness in overlapping and co-located cells, both in terms of average throughput and alternating access to the medium. On the downside, the proposed prioritization scheme causes a performance degradation compared to the standard Contention Window setting of about 20 to 30 percent in some of the scenarios. Considering the big advantage of the fairness improvement, such a tradeoff is rather acceptable.

Unfortunately, this approach can not be implemented in the case of a IEEE 802.11b network. The reason is that all stations associated with a single access point are treated alike. The access point can not distinguish the different stations. Therefore, these problems remain in pure DCF mode. However, in the following section we show how the IEEE 802.11e standard with its enhanced prioritization mechanisms can be used to implement the solution.

Voice Traffic performance

Section 3 introduced different voice codecs that are frequently used in voice applications. However, these codecs are very different in terms of bit rate, frame size, and look ahead size. Therefore, they differ in their suitability for Wireless LAN scenarios. In this section, we study the combination of Wireless LAN and voice clients using different voice codecs.

The simulation scenario is shown in Figure 5.40. A number of stations (Voice clients) is located in a single Wireless LAN cell. They perform voice conferences with clients on the wired network behind the access point using the Distributed Coordination Function. All wired components are connected over a 100 Mbps Ethernet link, so the delay on the wired network can be ignored.

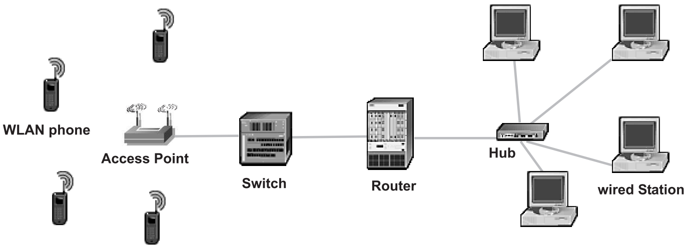

Figure 5.40: *Voice clients in a Wireless LAN environment*

Figure 5.41 shows the maximum number of clients in a Wireless LAN cell for variable frame times and different data rates. If a clients uses a frame time of 30 ms and a data rate of 5.3 Kbps (the default settings of the G.723.1 standard [IT96c]), it is possible to support 18 clients with an acceptable end-to-end delay and a delay variation of less than 1 ms. Due to retransmissions on the wireless network, no packet loss occurs on a higher layer in the scenario.

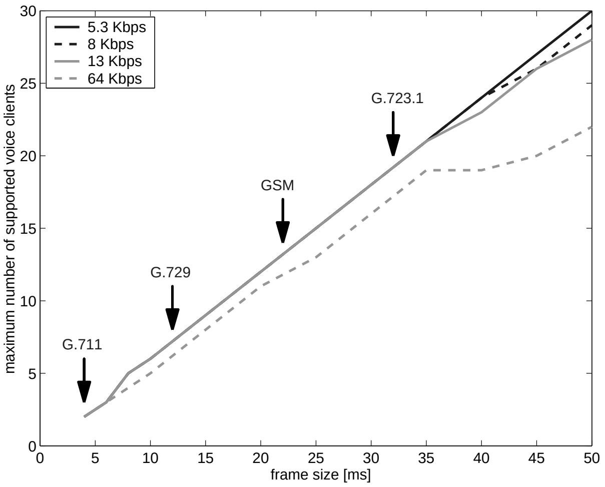

Figure 5.41: *Maximum number of voice clients in DCF mode*

If the GSM codec [Eur94] is used, only about 12 simultaneous voice calls can be supported. The G.729 standard [IT96b] only allows 6 concurrent voice clients. The most widely used G.711 standard [IT93a] performs worst in a Wireless LAN environment. Only up to 2 simultaneous voice calls can be supported.

Simulations with varying frame sizes and differing bit rates were performed. The results for the bit rates are shown as different lines. It can easily be seen, that not the data rate is responsible for the maximum number of clients. The main cause for the end-to-end delay is the frame size shown on the X-axis. The larger the frame size, meaning the larger the packet interarrival time of the voice codec, the more clients can be supported.

These results clearly show the importance of the voice codec on the performance of the Wireless LAN network. This fact has to be taken into account if a Wireless LAN environment should support voice conferencing. The conclusion has to be drawn that the only two voice codecs suitable for Wireless LAN networks are either G.723.1 or GSM. In the following, we constrict our simulations on the G.723.1 standard, since it shows the best results.

Similar results are found for the Point Coordination Function. However, since PCF is not suitable for multiple cell scenarios, we do not discuss the results here.

5.3.3 Overlapping and co-located cells in HCF mode

The last section presented a detailed analysis of the effect that overlapping and co-located cells have on the DCF operation mode in IEEE 802.11 networks. It turned out that fairness problems arise once there are overlapping areas. Depending on the type of overlap, a different prioritization scheme has to be applied in order to restore the fairness of the MAC protocol. The drawback is a reduced achievable throughput.

In this section, we want to extend the proposed solution to the IEEE 802.11e Hybrid Coordination Function MAC protocols. The goal is to support different priority classes, such that the fairness problem in overlapping and co-located cells can be solved. On the other hand, the HCF access control protocol is supposed to provide QoS capabilities to Wireless LAN networks. Therefore, service differentiation has to be provided by the MAC protocol even in the multiple cell scenarios. In this section, different QoS aware applications are simulated in order to evaluate the HCF protocols.

One of the conclusions drawn in the last section was that the prioritization scheme that is necessary to provide fairness in multiple cell scenarios, can not be implemented in IEEE 802.11b networks, since the DCF MAC protocol does not support service differentiation. All stations are treated in the same way. The HCF MAC protocol extensions is supposed solve this problem. One additional goal in this section is to clarify the question whether HCF is flexible enough to do this.

Traffic Model

As for DCF, the same simulation scenarios as described in Section 5.3.2 are studied for the HCF case. User mobility is still not considered in our simulation. The stations are located at the positions as specified in the scenarios. The goal here is to study the impact of the user positions on the performance of the Wireless LAN MAC protocol. Mobility is the topic of later sections. The Wireless LAN clients use the Hybrid Coordination Function (HCF) MAC protocol of the IEEE 802.11e standard [IEE03d]. HCF is explained in detail in Section 5.2.5. It provides different priorities in order to support service differentiation to QoS aware applications. In HCF, voice applications are supplied with the highest priority. The next highest level is applied to video transmissions, while the background FTP traffic always gets the lowest priority.

In order to evaluate the prioritization mechanism of the IEEE 802.11e standard in multiple cell scenarios, all three different traffic types are considered. They are discussed in detail in Section 3. Voice traffic as in interactive, bi-directional voice calls with different voice codes is the top-level priority application. The next highest priority usually goes to the video traffic, such as video conferencing. Finally, non-prioritized background traffic has to be considered as well. Again we consider FTP traffic as a worst-case scenario of Web traffic.

In all our simulations, the effect of the backbone network on the performance is not under study. Therefore, the backbone network is ignored where possible. The access points usually act as the terminating endpoint of the data communication in our simulation environment. The results are presented in three different parts. The first part considers the overlapping cells scenarios, while the second part is confined to the co-located cells scenarios. In the third part a combined solution is presented.

Overlapping cells and HCF

First, we consider the overlapping cells scenario B from Figure 5.25. It is asymmetric, since only station C2 is in the overlap of both access points, while station C1 is not disturbed by the data transmission of the access point A2. The worst-case scenario here is that the station C2 performs a QoS demanding application, while station C1 downloads files from the access point A1. This is due to the fact, that in an overlapping cell scenario the client within the overlap, here client C2, is disadvantaged as shown in Section 5.3.2. In the following, the station C1 performs 1 MByte file downloads. The goal is to adapt the IEEE 802.11e access machanisms such that the station C2 can perform real-time applications even if it is disadvantaged over station C1.

In the case of voice traffic and standard DCF or HCF operation, the MAC protocol can not provide an acceptable VoIP service for station C2. In DCF mode, the mean packet loss for the voice client C2 reaches 59.97 percent, which maps to a MOS score of 1.0 meaning *not recommended*. In HCF mode the average packet loss for station C2 even reaches 63.54 percent and again a MOS score of 1.0.

This is clearly not acceptable. DCF cannot provide any QoS, such that the results for DCF mode are not surprising. However, HCF with standard parameters already applies a much higher priority to the voice client than to the best-effort user. The problem is that with the standard parameters of $CW_{min} = 7$ and $CW_{max} = 15$, the collision probability is very high, since the retransmission attempts are performed after a rather short backoff period. Therefore, we can conclude that choosing such small Contention Window parameters is not suitable for the overlapping cells scenario B.

In order to overcome these problems, we adapt the Contention Window parameters as shown in Table 5.6. The set of priority classes is defined according to different CW_{min} and CW_{max} values. Table 5.10 shows the results for the new Contention Window parameters. For completeness, the results with the default DCF and HCF modes are shown as well.

MAC Protocol	Priority Class C1	Priority Class C2	Packet Loss C2 [%]	MOS Score
DCF	default	default	59.97	1.0
HCF	default	default	63.54	1.0
HCF	4	1	7.64	< 2.6
HCF	4	2	8.29	< 2.6
HCF	5	1	0.53	3.428
HCF	5	2	0.77	3.371
HCF	6	1	0.00	3.704
HCF	6	2	0.04	> 3.6
HCF	6	3	0.03	> 3.6
HCF	6	4	0.03	> 3.6
HCF	6	5	0.39	3.535

Table 5.10: *Overlapping cells scenario B: MOS values (1 MByte FTP files)*

An acceptable solution for this problem can be found when applying the priority classes (5,X) or (6,Y) with $X \in \{1, 2\}$ and $Y \in \{1, 2, 3, 4, 5\}$. In case of priority class (4,1) and (4,2), the MOS lies below 2.6 and leads to a user satisfaction which is *not recommended*. For priority classes (6,1), (6,2), (6,3), and (6,4) the voice quality is still *acceptable* with just a few users dissatisfied. For priority classes (5,1), (5,2), and (6,5) the voice quality drops just below *acceptable*.

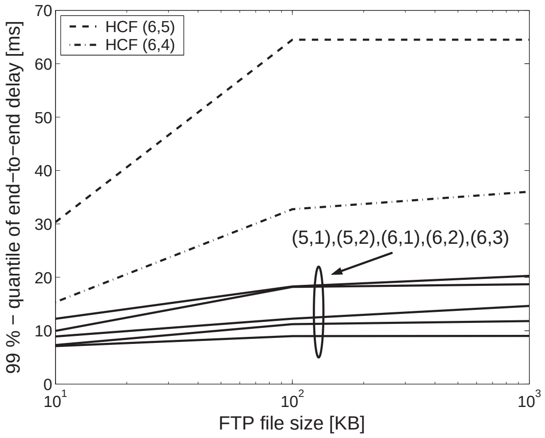

Figure 5.42: *Overlapping cells scenario B: voice delay*

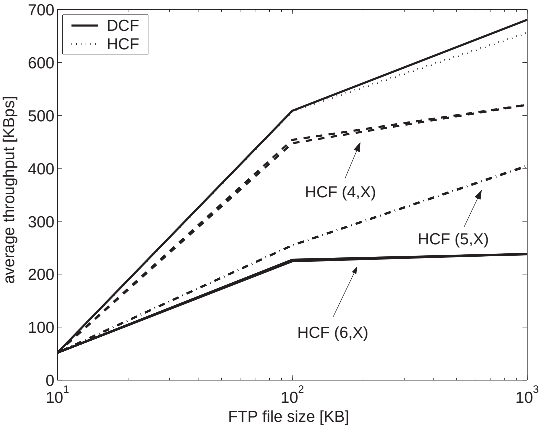

Figure 5.43: *Overlapping cells scenario B: FTP throughput (voice)*

The results show that station C1 must at least have priority class 5. These results are summarized in Figures 5.42 and 5.43. The 99 percent-quantile of the end-to-end delay of the voice application is shown in Fig. 5.42. It gives us proof that in scenarios with varying FTP load (depending on the FTP file size), the results that were described above still hold. One drawback of lowering the priority setting of the best-effort FTP user can be seen in Fig. 5.43. It shows the average throughput in KBps that the FTP user experiences. Clearly, the lower the priority (larger value means lower priority), the lower the average throughput gets.

However, as it is more important to provide QoS service than maximum throughput in the Wireless LAN scenarios considered here, choosing a priority setting of (5,X) is a good compromise. A good FTP performance can still be reached without interfering with the voice application.

MAC Protocol	Priority Class C1	Priority Class C2	Packet Loss C2 [%]	PSNR	MOS Score
HCF	default	default	86,19	12.66	Bad
HCF	4	1	6.84	25.45	Fair
HCF	5	1	5.67	25.69	Fair
HCF	5	2	6.01	24.60	Fair
HCF	5	3	6.34	26.55	Fair
HCF	6	1	0.07	40.67	Excellent
HCF	6	2	0.19	40.97	Excellent
HCF	6	3	0.43	46.84	Excellent
HCF	6	4	0.53	45.27	Excellent

Table 5.11: *Overlapping cells scenario B: PSNR values (1 MByte files)*

Table 5.11 shows the results for the case of video traffic. It can be seen that priority classes (4,1) and (5,X) with $X \in \{1,2,3\}$ only provide *fair* video quality (MOS=3). If the priority set (6,Y) with $Y \in \{1,2,3,4\}$ is used, the MOS value changes to 5 indicating *excellent* video quality. The

PSNR is always above 37 in all simulation runs.

Again, Figure 5.44 shows the 99 percent-quantile of the end-to-end delay in ms for the video applications. Figure 5.45 depicts the average throughput the FTP user experiences when applying different priorities.

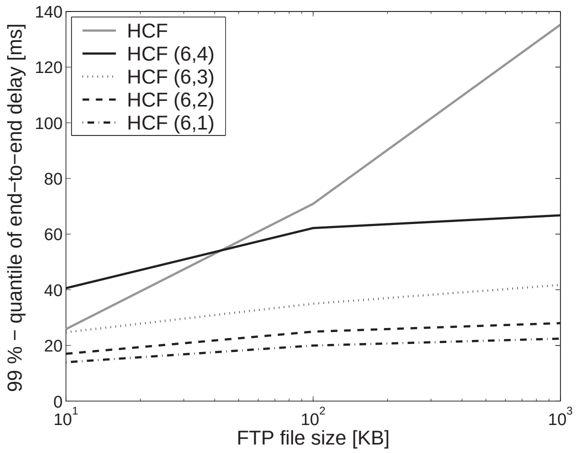

Figure 5.44: *Overlapping cells scenario B: video delay*

Thus, for overlapping cells B we can conclude that when applying different priority settings to the voice, video, and best-effort user, it is possible to provide QoS and still allow FTP users to get good throughput rates. Different priority settings are possible and can be used by WLAN Internet Service Providers to adapt the settings to specific needs.

Overlapping cells A is the only symmetric overlapping cells scenario. Both stations are located in the overlap and both experience problems in the case of default HCF parameters. However, since both stations experience the same problems, the solution is easier than in the former case of overlapping cells B. Here, the priority settings (3,1) and (4,1) are

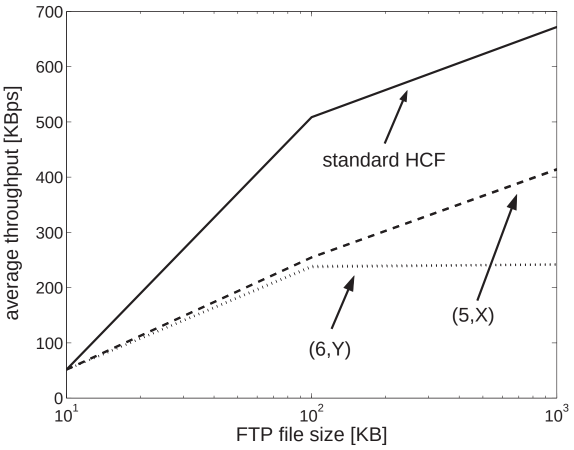

Figure 5.45: *Overlapping cells scenario B: FTP throughput (voice)*

already sufficient. This means that the priority of the FTP user can be higher here, compared to the former case. This allows the FTP station to receive an even higher share of the bandwidth than before.

Overlapping cells C, on the other hand, behaves almost exactly like overlapping cells B. The results are shown in Figures 5.46 and 5.47, while the results for the video case can be seen in Figures 5.48 and 5.49.

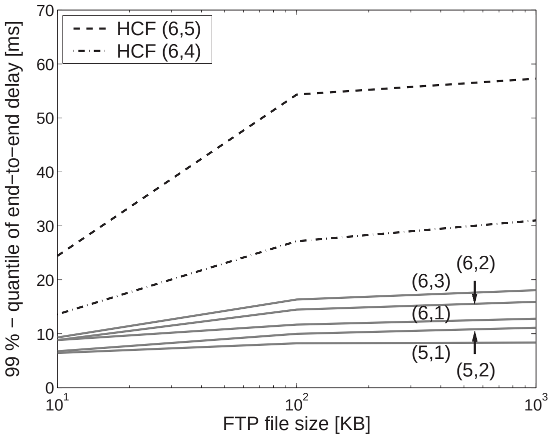

Figure 5.46: *Overlapping cells scenario C: voice delay*

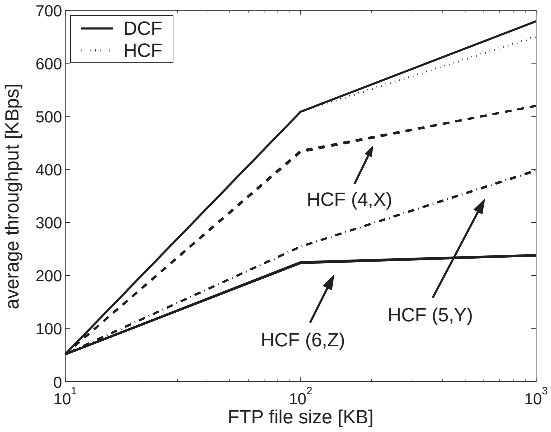

Figure 5.47: *Overlapping cells scenario C: FTP throughput (voice)*

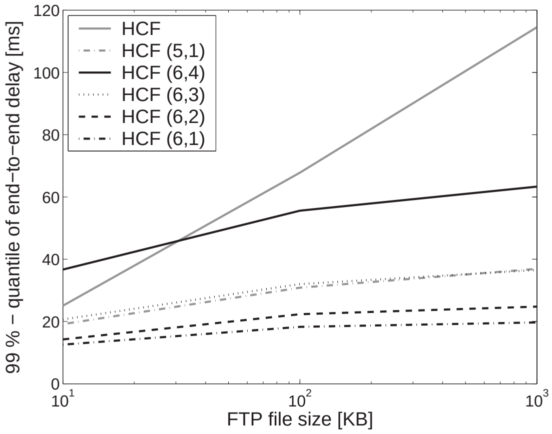

Figure 5.48: *Overlapping cells scenario C: video delay*

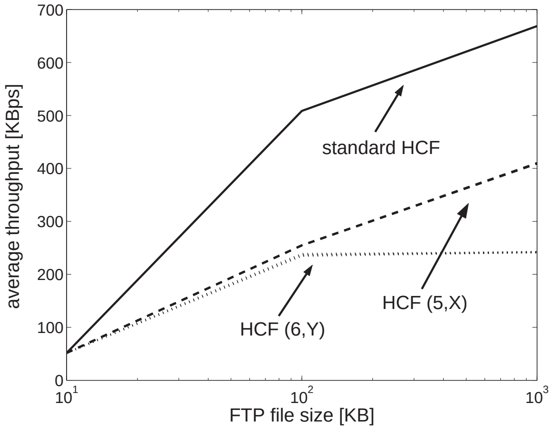

Figure 5.49: *Overlapping cells scenario C: FTP throughput (video)*

Again, we can conclude that there exists a number of different priority settings that can be used in order to provide QoS. The priority of the FTP clients has to be set low enough in order to not disturb the QoS demanding application. On the other hand, it should be as high as possible in order to allow a maximum throughput.

Co-located cells and HCF

In the case of co-located cells, the situation turns out to be less critical. Due to the fact, that the access points are within the reception range of each other, the transmitting FTP server is interfered by at least one QoS client (in our case the other access point). Therefore, the standard HCF parameters prove to be sufficient in these cases.

The 99 percent-quantile of the end-to-end delay is 10 ms. The MOS value for voice is always above 3.6 indicating *acceptable* quality. For the video client, the 99 percent-quantile of the end-to-end delay is less than 10 ms, no packets are lost, such that the video quality is *excellent*. The FTP performance in the case of standard HCF parameters is as good as it can get.

Combined Solution

The goal of the QoS enabled MAC protocols is to provide QoS for voice and video applications at the same time. In order to evaluate our priority settings for such a case, we simulated the worst-case scenario, overlapping cells B, with station C2 using voice and video at the same time. Station C1 still performs FTP downloads. The priority settings (6,2,1) are chosen. This means that the voice application uses priority class 1, the video application was configured to use priority class 2, while the best-effort FTP traffic was handled with priority class 6. The results are shown in Table 5.12.

Traffic Type	Prio. Class	Packet Loss [%]	Delay 99%-quant. [ms]	Delay Max.	Jitter [ms]	MOS Score
Voice	1	0.03	10.77	22.12	5.68	>3.6
Video	2	0.27	34.94	59.53	75.04	5

Table 5.12: *Combined solution (1 MB files), voice and video*

It can be seen that HCF with priority class (6,2,1) can provide adequate QoS even if both multimedia applications are used in a single station. The same simulation with default HCF parameters results in packet loss for both voice and video applications above 80 percent, which certainly provides bad voice and video quality. The best-effort FTP user suffers a performance degradation in terms of average throughput of about 50 to 60 percent.

Summarizing the simulation results found in this section, we can conclude that the prioritization parameters as proposed for the HCF operation mode, are not sufficient in overlapping and co-located cells. They can prioritize certain stations, but they lead to high levels of packet loss, and thus to large quality degradation in case of voice and video applications. Our studies also showed that different sets of prioritization parameters can be applied that provide the required level of prioritization while still allowing high medium utilization.

QoS support in large-scale Wireless LAN environments is possible (see also [PHWTG04], [GK03], [MCM$^+$02]). However, further studies should focus on additional mechanisms that are necessary for Wireless LAN to become a 4G technology. One such issue is admission control. It is only possible to support QoS in any (wireless) network, if the number of users is kept below a certain threshold. This number is influenced by the type of traffic that is used and the environment considered, e.g. overlapping or co-located cells.

One other issue in wireless networks is handover techniques, especially in QoS demanding scenarios. It has to be clarified if the handover mechanisms that have been defined for Wireless LAN can be used while still keeping the performance at the necessary level. We focus on this issue in the remainder of this chapter.

5.4 Handover Mechanisms

Wireless LAN was introduced as an extension of wired local area networks, e.g. in locations where cables are not an option. However, the huge success of Wireless LAN shows that the users demand for the great flexibility and mobility that a wireless network can provide them with. Whole office buildings and campuses have been covered by a single Wireless LAN network in the meantime. The user can wander about the covered area and seamlessly stay connected to the network all times.

On the network side, it has to be made sure that the connection of the users is handed over from one access point to another, once the coverage areas are crossed. Such handover procedures are necessary in all large-scale wireless networks.

Several different approaches to perform a handover are known from legacy mobile networks. The easiest is the hard handover, also known as *break-before-make*. The connection to the "old" base station is disconnected before the connection to the "new" base station is initiated. There is always a certain delay in which the mobile is not connected during the handover process. The goal of a hard handover is to keep this delay as short as possible.

More sophisticated handover mechanisms have been introduced in CDMA based mobile networks, such as soft and softer handover (see [Rap96], [Sch03], [Tyc02]). With soft handover a mobile sets up a number of connections to all the base stations within a certain range. Therefore, a handover does not imply a delay in which the mobile is not connected.

On the contrary, these multiple connections can even be used to lower the bit and packet error probability as shown in [HSL02]. Softer handover is an extension of soft handover. It defines that a mobile can have several connections to the different sector antennas of a single base station. This makes sure that the handover between different sectors is "soft" as well.

A Wireless LAN, on the other hand, does not allow such highly sophisticated techniques. The technology is supposed to be cheap. Therefore, Wireless LAN only allows to perform hard handovers between the different access points. However, this still allows a number of different approaches, that greatly differ in terms of performance [MSA02]. In this section we study the different proposed handover techniques. The goal is to find a handover mechanism that allows handovers that keep the delay short enough to support voice calls and video conferences.

A hard handover in Wireless LANs consists of three different parts. First, scanning is performed by a roaming station. Scanning is used to find the access points that surround a station and to perform the handover decision. The second part is authentication. It is used to make sure that the station is allowed to access the network. Finally, the association or reassociation takes place. It is the actual connection setup phase. In the following we discuss these three parts in greater detail.

5.4.1 Scanning

The basic parameter for a roaming station is the *Signal-to-Noise Ratio (SNR)*. As soon as the SNR drops below a certain threshold, called the *Cell Search Threshold*, the station starts the handover process. During this process, the station searches for new access points and if the difference between the SNR of the old access point and the potential new access point reaches a threshold known as the *Delta SNR*, the station initiates the actual handover [Luc98]. This process in shown in Figure 5.50.

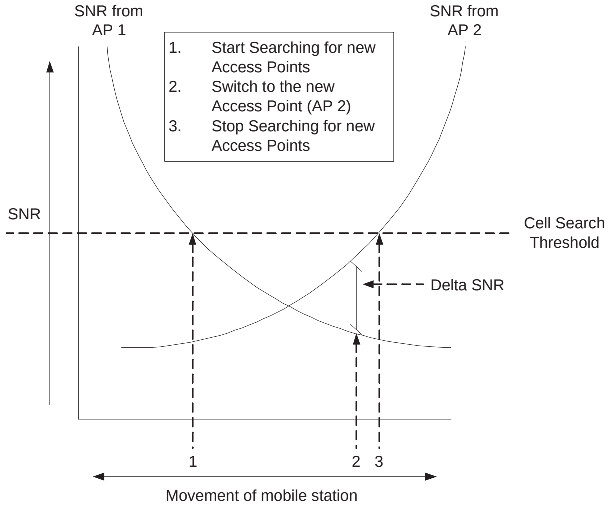

Figure 5.50: *Signal-to-Noise Ratio and handover decision*

As the station moves away from the access point, from left to right in the figure, the SNR from access point 1 (AP1) decreases. At the same time the SNR from access point 2 (AP2) increases as the station moves closer. Once the SNR of AP1 drops below the Cell Search Threshold, indicated by position 1 on the horizontal axis, the roaming station enters the Cell Search state and starts its scanning process. When the station moves closer to the second access point, the difference between both SNRs finally exceeds the Delta SNR value and the station switches over to the second access point (position 2), but remains in the Cell Search mode until the SNR has passed the Cell Search Threshold, marked as position 3.

If the SNR keeps decreasing as the station moves farther away from the

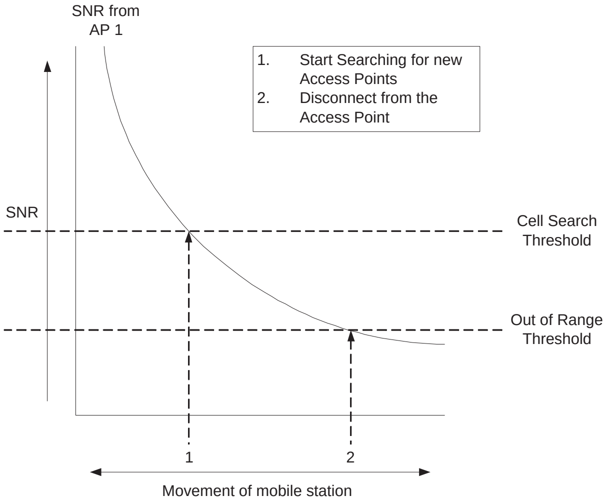
SNR from AP 1
1. Start Searching for new Access Points
2. Disconnect from the Access Point
SNR
Cell Search Threshold
Out of Range Threshold
1
2
Movement of mobile station

Figure 5.51: *Signal-to-Noise Ratio and station out of range*

access point while no other access point can be found, the SNR ultimately drops below another threshold, called *Out of Range Threshold*, as shown in Figure 5.51. In a situation where the roaming station experiences a SNR below the Out of Range Threshold, it may fall back in speed from 11 Mbps to 5.5 Mbps or even down to 1 Mbps. If the SNR still is below the Out of Range Threshold, the station disconnects from the access point.

The different thresholds can be chosen depending on the access point density parameter. The access point density can be set by the user or administrator. Three different values are allowed: low, medium, and high. Table 5.13 shows the Cell Search Threshold and the Delta SNR value for these densities.

Threshold	Access point density		
	Low	Medium	High
Cell Search [dB]	10	23	30
Delta SNR [dB]	6	7	8

Table 5.13: *Access point density parameters*

They are used to avoid *handover oscillation*, a situation where the station performs rapid handovers between different access points. The higher the access point density, the higher the values of Cell Search and Delta SNR Threshold are chosen. Thus, the coverage area of a single access point is smaller for the higher densities, while the difference between the old and the new access point has to be larger before a handover is performed. Therefore, fewer candidates for a potential handover are found and the number of handovers can be reduced.

As already explained above, the scanning procedure is started, once a potential candidate is found. In the following, the different scanning mechanisms are introduced. In principle, two different scanning mechanism can be distinguished: *active* and *passive* scanning. Passive scanning is performed by listening to the medium for networks to announce themselves. With active scanning, the station takes on an active role and transmits special packets to identify the environment that it is currently located in.

The scanning process is very important as it is the first step of a handover. The two simple approaches still leave room for further extensions. One such extension is the Neighborhood Detection [JWKZ03]. Although it has not yet found its way into the official Wireless LAN standard, it is discussed in this section as well.

Passive Scanning

In the Passive Scanning mode, a station does not send any data on the wireless medium. It scans the medium for networks to announce themselves. Each access point transmits Beacon frames for this purpose. The station listens to each channel for a specific time interval and waits for the Beacons. The received Beacons are buffered by the station and the important information is extracted. Beacons are designed to allow a station to receive the necessary information about the Basic Service Set of the access point. It contains information about the supported data rates, the Service Set ID, the physical layer characteristics, and information about the supported MAC protocols. Two different types of passive scanning can be distinguished, the *normal passive scanning* and the *fast passive scanning*.

Normal passive scanning defines that a station switches to the first channel allowed by the regulatory domain and waits for Beacon frames. If no Beacon is received after a specific time, the station switches to the next channel until all channels are scanned. The station scans each channel which results in a great overhead, because it receives the Beacon frames from all the access points that use the scanned or one of its overlapping channels.

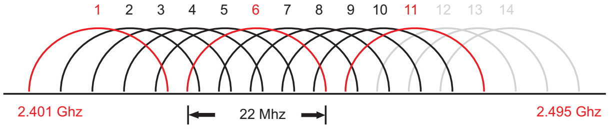

Figure 5.52: *Overlapping channels in the Wireless LAN 2.4 GHz band*

Fast passive scanning, on the other hand, exploits the overlapping channels of Wireless LAN as shown in Figure 5.52 for the 2.4 GHz band. The station only scans the non-overlapping channels, e.g. the channels one, six, and eleven. This fast passive mode reduces the period of time used for scanning compared to the normal passive scanning. However, it is more error-prone. The signal of an access point operating in an overlapping channel is received with a smaller SNR, which has to be accounted for. On the other hand, if a Beacon frame is delayed due to a previous data transmission, the scanning station could switch to the next channel without receiving a Beacon frame. If all the channels were to be scanned, such a scenario becomes less probable, since there are several scanning periods in which the Beacon of a single access point can be received.

Active Scanning

A more active role is designated to the stations when active scanning is performed. They do not just listen to the traffic on the different channels, but they actively try to find potential access points by transmitting Probe frames.

First, the channel is passively scanned for a specific period of time, the Probe Delay Time. If the channel is found to be in use, i.e. packets are transmitted, the scanning procedure probes this channel. This timeout is used to keep an empty channel from blocking the remaining scanning procedure. Once the station decides to actively scan the channel it transmits a Probe Request frame using the normal DCF mode. If the channel stays idle for a time referred to as MinChannelTime after the Probe Request frame was sent out, the station moves on to the next channel. If, however, the channel gets busy within the timeout period, the MinChannelTime timeout is canceled and the station continues to listen for Probe Response frames until the maximum time, called MaxChannelTime, has passed.

An access point answers a Probe Request frame with a Probe Response frame. The station in turn acknowledges the successful reception

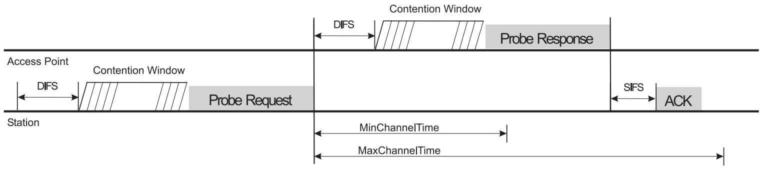

Figure 5.53: *Active Scanning procedure for a single channel*

of the Probe Response frame. This is necessary, since the Probe Response frames are sent in DCF mode as well. If the successful reception is not notified to the access point, it tries to retransmit the packet as usual. The whole procedure is depicted in Figure 5.53.

A Probe Request frame is answered by all the access points in the reception range of the station. However, if there are several access points trying to transmit Probe Response frames using DCF, the transmission of some of the access points is delayed due to the transmission of the other access points. Therefore, the higher the access point density, the larger the MaxChannelTime needs to be in order for all access points to be able to answer the station's call. This topic is evaluated later in this section.

The original IEEE 802.11 standard only defines the normal scanning operation, where a station performs the scanning procedure on each of the channels allowed by the local regulatory domain. Nevertheless, as for the passive scanning mechanism all access points even those operating in overlapping channels, receive the Probe Request and send a Probe Response. Therefore, in normal active scanning each access point is found repeatedly. This again allows to optimize the scanning procedure. In recent years, a number of different active scanning mechanisms have been proposed. The two most promising approaches are *fast active scanning* and *active scanning with neighborhood detection.*

The fast active scanning mechanism simply utilizes the fact that all access points operating on the scanned or one of the overlapping channels announce themselves by transmitting a Probe Response frame. Rather than scanning all the channels of the regulatory domain, only non-overlapping channels have to be scanned. As for fast passive scanning this could be channels one, six, and eleven. This helps to lower the overhead of the scanning operation. On the downside, the same problems arise as for fast passive scanning. One problem is that the SNR measured by the station has to be adjusted if the access point operates in an adjacent channel. The other problem is that an access point's Probe Response frame could get lost a couple of times in normal active scanning mode without disturbing the operation. This is caused by the fact that an access point has several chances to announce itself to a calling station. In fast active scanning mode, an access point usually has just one single chance.

Active Scanning with Neighborhood Detection

The two active scanning mechanisms introduced earlier, do not have any information about the network they are located in, once they start scanning. A number of channels has to be scanned and all access points in the reception range of the scanning station are supposed to announce themselves. However, since scanning is performed every time a handover should be initiated, it is crucial to the network to speed up the operation as much as possible.

One promising active scanning approach discussed in the literature is active scanning with neighborhood detection. It is introduced in [JWKZ03], [BS03], and [PC02]. The main idea is that the access points do not only announce information about themselves in the Beacon frames, but they also include information about neighboring access points. This allows all associated stations to directly communicate with all potential neighboring access points. Probe Response frames also include the neighborhood information. The maximum number of neighboring access points that can be announced that way is set to twelve.

It is left to the access points to retrieve the neighborhood information from their (sub-)network. In the easiest case, all the access points are located in a single subnet, such that discovery is an easy task. However, if there are several different subnets involved, more sophisticated techniques have to be applied. Another important issue is, what access points are to be announced if there are more than twelve candidates available. In [PC02], the list of neighboring access points is calculated using handover probabilities for specific access points with the movement ratio. Alternative approaches are discussed as well.

The actual scanning mechanism can be performed in different ways. One way is shown in Figure 5.54. The scanning station transmits a Probe Request frame to one of the potential access points using normal DCF operation. In turn, the access point answers the call by issuing a Probe Response frame after a SIFS interframe space. This is called fast active scanning with neighborhood detection. Here, the term fast refers to the instantaneous answer of the access point. Such an approach becomes possible, since the scanning station sends its request directly to the access point, rather than issuing a broadcast.

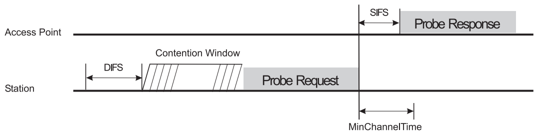

Figure 5.54: *Fast Neighborhood Scanning procedure*

On the contrary, the access point might send its Probe Response frame at a later time as shown in Figure 5.55. This normal neighborhood scanning operation specifies that the access point acknowledges the reception of the Probe Request frame with a normal ACK packet. After that it responds with the Probe Response frame using either a PIFS interframe space or normal DCF operation with a standard backoff.

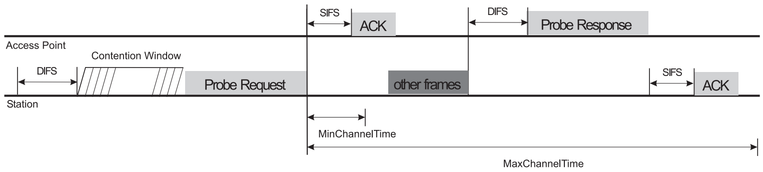

Figure 5.55: *Normal Neighborhood Scanning procedure*

These different approaches for the neighborhood scanning procedure are discussed in more detail in [JWKZ03]. In the following, only fast neighborhood scanning is studied, since it clearly defines the fastest scanning mechanism.

5.4.2 Authentication

The second step of a handover that a station has to perform is the authentication at the new access point. The goal of the scanning procedure as described above is to find the best candidate. Once the station has decided which access point it wants to roam to, the authentication procedure in the access point decides whether the station is allowed to associate to it.

Two different authentication procedures are defined in the initial IEEE 802.11 standard. The *open system authentication* is not really an authentication process, since any station is by default allowed to access the network. It is merely performed as a type of a placeholder for more advanced authentication mechanisms.

The more widespread type of authentication is the *shared key authentication*. It uses the Wired Equivalent Privacy (WEP) algorithms to decide on the access request by the station [PF03]. Here, a shared secret key is used for the purpose of authentication. The access point checks if the station has been configured with the same shared secret key that has been configured within the access point's configuration. Usually, the network administrator is responsible for the management of these secret keys. Once the access point positively checks for the right secret key, the station receives access.

One other potential authentication mechanism is described in [PC02], referred to as *preauthentication*. It is based on the fact that the standard does not necessarily demand an explicit authentication of a station at each access point that it tries to connect to. Therefore, an authentication could be performed the first time that a station enters a network, which is valid not only for this single access point but for all access points within this network. Such a preauthentication allows a station to skip the authentication procedure in case of a handover within a single network.

Other types of preauthentication are discussed in the literature as well. In [Gas02], the authors propose that a station authenticates with several access points during the scanning process. Whenever one of these access points is entered, the authentication procedure can be skipped.

Authentication is a part of the Wireless LAN security, which is identified as one of the major drawbacks in today's legacy systems. Many of these security mechanism have been shown to be weak. Wireless LANs operated within a company, however, necessitate a high level of security. This should be considered an important part of the planning process.

5.4.3 Association and Reassociation

The previous sections explained the scanning and authentication procedures. If both of these procedures are finished successfully, then the association or reassociation is the only task that remains before the station is connected to the network and the handover is completed. The goal of this task is to register the station on the (wired) network, such that all data traffic destined to the station is sent through the access point the station is currently associated with.

This can for example be achieved by sending an ARP message within the wired backbone, such that the MAC address of the station is associated with the switch port leading to the access point.

Two different situations can, however, be distinguished. In case the station newly entered the Wireless LAN network and, thus has not been connected so far, it has to associate itself with the access point. A handover represents the second case. The station has already been associated to one of the access points in the network. In order to perform a handover, the reassociation takes place. In addition to an association, the reassociation has to make sure that the station's association at the old access point is deleted. The Wireless LAN standard explicitly forbids multiple associations of a single station.

Association

A station that newly enters a Wireless LAN network has to perform an association procedure after successful authentication. For that purpose, the station sends an Association Request frame to the access point. This Association Request contains information about capabilities, listen interval, SSID, and supported data rates of the station. The capability information is used for PCF operation. It tells the access point if the station should be polled. The listen interval is used to inform the access point about the times the station wakes up to listen for Beacon frames. This is necessary if the station is run in power saving mode.

The access point acknowledges the successful reception of the Association Request by issuing an Association Response frame, which informs the station about PCF capabilities, status of the association, Association ID, and supported data rates of the access point. The PCF capabilities are used to indicate whether the access point acts as a Point Coordinator or not. The status of the association shows the station that the association has been competed successfully or otherwise what the reason for the failed Association Request is. The Association ID is merely used internally to identify the station. The station acknowledges the Association Response frame which completes the association procedure of the station.

In case of a successful association, the access point informs the wired network of the new station by transmitting an Address Resolution Protocol (ARP) packet within the wired backbone. This configures the network to forward the packets destined for the station to its associated access point.

As pointed out earlier, there are cases where the other access points within the (local) network have to be informed about the new station. One such case is preauthentication. The access points are equipped with the Inter Access Point Protocol (IAPP) (see [IEE03a]) for this purpose. It is explained in greater detail in the following section.

Reassociation

In case of a handover within an Extended Service Set (ESS), the station is already associated to some access point within the network. Therefore, the normal association procedure is not suitable to perform such a handover, but a reassociation has to be issued. On the wireless side, a reassociation is almost identical to the association, but on the wired backbone side, more data has to be exchanged. The reassociation procedure is shown in Figure 5.56.

The reassociation is used to inform the old access point of the new location of the station. In addition to the information within an Associ-

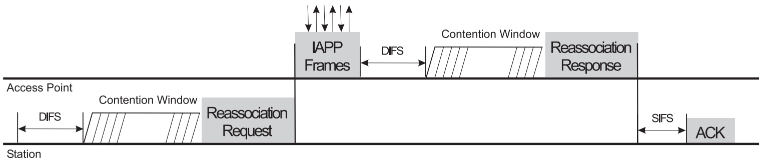

Figure 5.56: *Reassociation procedure*

ation frame described in the last section, a Reassociation frame contains the MAC address of the old access point. The new access point, therefore, can use the Inter Access Point Protocol to validate the request as explained in more detail in the next section. Once the new access point has validated the information of the station, it sends a Reassociation Response frame to the station, which is acknowledged by the station.

After the successful transmission of the Reassociation Response frame, the new access point is responsible for the newly associated station. The old access point deletes the association and forwards the buffered packets for the station to the new access point.

It is worth mentioning that there are proposals for a fast reassociation procedure in PCF mode. It is presented in [GW99]. Rather than using the DCF mode and CDMA/CA operation, polling is used by an access point to search for new stations within its reception range. The access point transmits *Who is New (WN)* frames in the Contention-free Period, which inform the stations about the correctly associated stations. Then, a (Re)Association interval follows, which allows stations to transmit association requests to the access point.

Nevertheless, since polling mechanisms have major problems in the multiple cell scenarios considered in this work, PCF and the fast reassociation procedure can not be considered a good candidate and are, therefore, not considered in the remainder of this work.

Inter Access Point Protocol (IAPP)

In large-scale Wireless LAN implementations, where handover between the involved access points occur frequently, there is a need to exchange information between the access points. Therefore, the IEEE standard 802.11f [IEE03a] was defined. It specifies the Inter Access Point Protocol (IAPP) that allows communication between the access points of a single ESS. The IAPP is located on top of the Wireless LAN MAC protocols as the stack in Figure 5.57 shows.

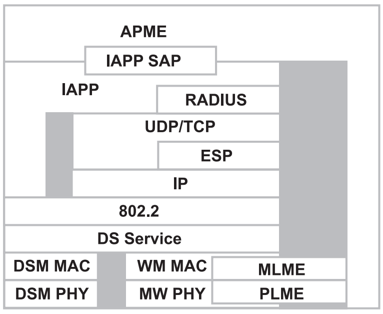

Figure 5.57: *Wireless LAN protocol stack*

The IAPP protocol uses TCP for the communication between the access points, while UDP is used for the communication to the Remote Authentication Dialing User Service (RADIUS), which is used for centralized AAA management. The IAPP also has to make sure that the forwarding tables of switches or routers in the wired backbone network are updated in case of a handover. Therefore, IAPP can directly access the 802.2 layer in order to submit layer 2 update frames.

On top of the IAPP layer sits the Access Point Management Entity (APME). It utilizes all the services offered by the IAPP through the

Inter Access Point Protocol Service Access Point (IAPP SAP). The gray areas in the figure indicate that there is no communication between these layers. It is used to show for example that the APME layer does not communicate with most of the lower layers except the MAC Layer Management Entity (MLME) and the Physical Layer Management Entity (PLME) sublayers. These two sublayers allow the APME to configure the parameters of the MAC and Physical layers in order to adapt them to certain situations.

It was mentioned earlier that the IAPP is essential in cases where handovers are performed within a Wireless LAN network. They have to perform important administrative tasks. In the following the IAPP functionality in case of an association and reassociation is explained.

In case of an association, there is actually little the IAPP has to do. A station tries to newly get access to the network. However, the station could still be associated with some other access point. In such a case, the station should perform a handover and the reassociation procedure instead, but erroneous station devices have to be taken into account.

Figure 5.58 shows the data flow once an association procedure is initiated within an access point. The local MLME sublayer sends an indication to the local APME sublayer, telling it that a new station tries to associate. The APME issues an add request (IAPP-ADD.request) at the IAPP layer. The IAPP now transmits the layer 2 update frame on the wired backbone network. This updates all the forwarding tables of the devices in the local subnet, such that the packets destined for the station are forwarded through the access point.

Then, the local IAPP layer transmits a notify packet to all the other access points on the local network segment. This packet is used to inform all other access points of the ESS about the new station. All receiving access points check their local association tables for this station, and if they find an entry, the station is disassociated from this access point.

Finally, the local IAPP layer confirms the operation by issuing a con-

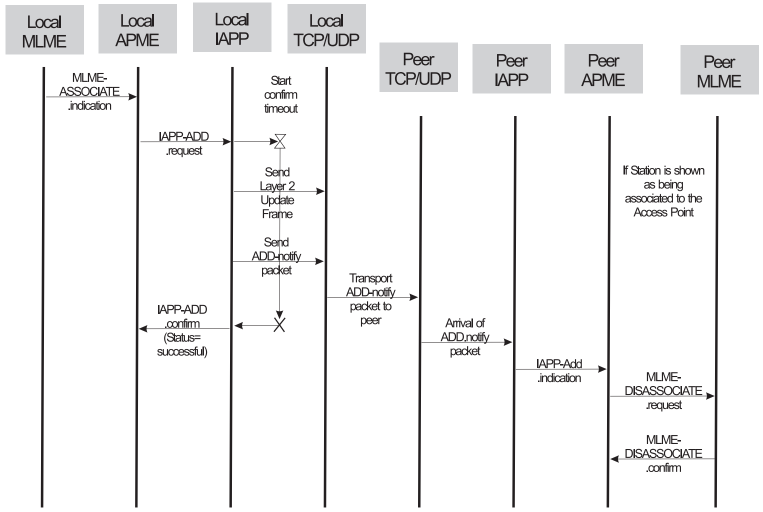
Local MLME
Local APME
Local IAPP
Local TCP/UDP
Peer TCP/UDP
Peer IAPP
Peer APME
Peer MLME
MLME-ASSOCIATE .indication
Start confirm timeout
IAPP-ADD .request
Send Layer 2 Update Frame
Send ADD-notify packet
If Station is shown as being associated to the Access Point
Transport ADD-notify packet to peer
IAPP-ADD .confirm (Status= successful)
Arrival of ADD.notify packet
IAPP-Add .indication
MLME-DISASSOCIATE .request
MLME-DISASSOCIATE .confirm

Figure 5.58: *Data flow during an IAPP Association procedure*

firm message to the local APME layer. This concludes the IAPP procedures in case of an associating station.

In case of a reassociation, the IAPP procedure is more complex. As shown in Figure 5.59 the operation starts with the MLME sublayer indicating the APME sublayer that a station tries to reassociate. The Reassociation Request frame from the station includes the information about the access point that it was formerly associated with. It is left to the IAPP layer to inform this old access point of the handover. Therefore, the MLME also forwards this information to the APME layer. The local APME layer on its part notifies the local IAPP layer of the reassociation request and the old access point. The IAPP layer informs the old access point about the movement of the station.

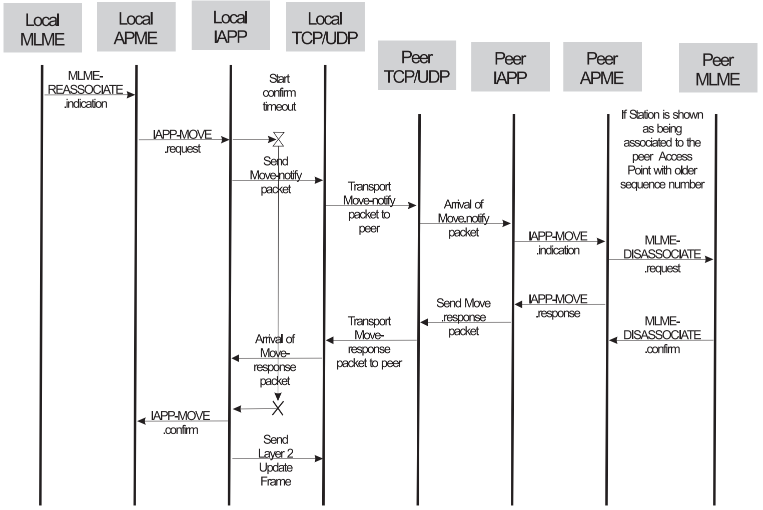
Local MLME
Local APME
Local IAPP
Local TCP/UDP
Peer TCP/UDP
Peer IAPP
Peer APME
Peer MLME
MLME-REASSOCIATE .indication
Start confirm timeout
IAPP-MOVE .request
Send Move-notify packet
Transport Move-notify packet to peer
Arrival of Move.notify packet
IAPP-MOVE .indication
If Station is shown as being associated to the peer Access Point with older sequence number
MLME-DISASSOCIATE .request
IAPP-MOVE .response
Send Move .response packet
MLME-DISASSOCIATE .confirm
Transport Move-response packet to peer
Arrival of Move-response packet
IAPP-MOVE .confirm
Send Layer 2 Update Frame

Figure 5.59: *Data flow during an IAPP Reassociation procedure*

This Move-notify packet is transmitted using TCP in order to make the transmission reliable. On the side of the old access point, the information traverses the IAPP and the APME layer and is finally passed to the MLME layer which takes care of the disassociation of the station. In turn, the old access point replies with a Move-response packet that acknowledges the successful reception of the Move-notify packet. Ultimately, the local APME layer of the new access point receives the IAPP-MOVE.confirm packet, that finishes the procedure.

Once, the movement of the station is confirmed by the new access point, the layer 2 update packet can be sent in order to update the forwarding tables of the devices in the local subnet. The old access point also updates its forwarding table. If this old access point still has buffered frames for the station, it can forward them through the new access point. On the other hand, if it turns out that the moving station can not be found in the association table of the old access point, the association request at the new access point is denied and the new access point transmits a deauthentication frame to the station.

This procedure shows that a station has to be deregistered from the old access point before it can be associated with the new access point, since a station is not allowed to associate with more than one access point at a time.

5.5 Handover Performance

The last section has shown the three tasks, scanning, authentication, and (re)association, that have to be performed when a handover occurs. It was pointed out that there are several ways to perform the different procedures. In the following, the different tasks are analyzed in terms of performance. The goal is to find a handover mechanism that can support the strict QoS requirements of voice and video applications [PH04]. This is a necessity in 4G networks, where Wireless LAN is supposed to play an important role.

Scanning, as the first step of the handover procedure, is responsible to decide on which access point to roam to. The last section showed that there are a number of different approaches to search the channels for access points. All of the presented procedures are evaluated and compared to each other. Authentication is an important part of a handover, as well. However, preauthentication can be performed, i.e. the station's initial authentication is valid for the whole (local) subnet. Thus, a renewal of the authentication is not necessary in the cases considered here, such that the authentication is not critical for the whole handover procedure.

(Re)Authentication on the other hand, necessitates to transmit packets to a neighboring access point within the backbone network. This communication has to be performed before an access point is allowed to give the station access to the network. The handover is not performed before the reassociation has finished. Therefore, reassociation is important for the performance of the handover and has to be accounted for.

The evaluation of the handover procedure is divided into four different parts. First, an empty network is considered. No background traffic can defer the handover. This scenario helps to understand the basics of the different approaches. In the second part, voice clients are studied. The scenario consists of a number of stations that perform voice calls with fixed clients in the wired backbone network. The third scenario concentrates on video traffic. Finally, the last part of this section considers a mixture of stations using different applications. It answers the question of whether there are suitable handover procedures for Wireless LAN in 4G environments.

5.5.1 Handover Performance in Undisturbed Environments

In order to understand the different handover approaches and to study their effect on the system performance, a system scenario is considered in this section that is not disturbed by any additional traffic on the wireless channels. The scenario as shown in Figure 5.60 is analyzed.

It consists of two access points AP1 and AP2, which are connected to a Switch using a 100 Mbps Ethernet connection. The switch is itself connected to a router which leads the data packets to the destination workstation in the wired backbone. A single station moves back and forth between the two access points. The access points are placed 70 meters apart from each other. Once the station moves about 50 meters away from its associated access point it starts the handover procedure according to the SNR rules explained in Section 5.4.1.

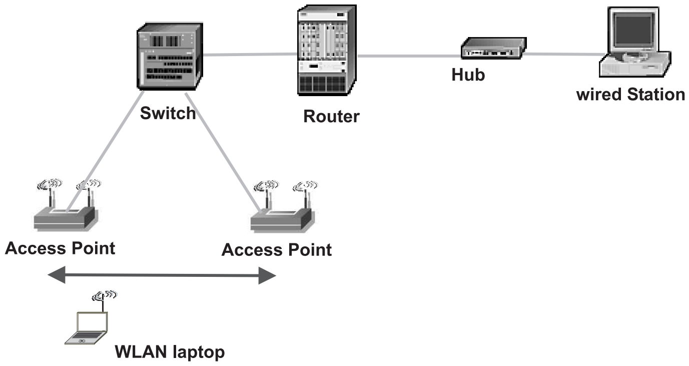

Figure 5.60: *Handover scenario without disturbing background traffic*

The station is preauthenticated to both access points, such that a renewed authentication in case of a handover is not necessary. The station assumes a low access point density, such that the scanning procedure is started once the SNR of the associated access point drops below 10 dB. The reassociation starts once the station finishes the scanning task and the new access point has been found. All scanning mechanisms are considered in the following.

In this simple scenario, the station does not perform any kind of application, i.e. no other than handover related traffic is transmitted. Therefore, the reassociation procedure, which consists of a number of packets being transmitted between the two access points, is not delayed, such that the reassociation task always amounts to approximately the same time. The differences in handover performance are, thus, solely caused by the scanning procedures.

Figure 5.61 shows the relation between the time spent for scanning versus the reassociation time. As just explained, the reassociation time

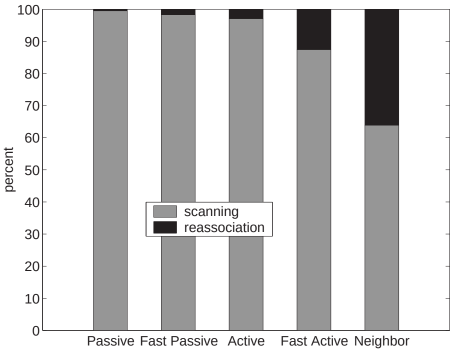

Figure 5.61: *Ratio between scanning versus reassociation delay*

can be assumed to be identical in all five cases. Therefore, the figure shows that scanning is by far the dominating factor. From about 60 percent to more than 98 percent of the total handover time is needed for the scanning mechanism. As expected the normal passive scanning performs worst, while the most advanced mechanism, Neighborhood scanning, is the fastest.

It is easy to see from these results, that scanning plays the most important role in a handover in terms of performance. Therefore, the different scanning mechanisms are explicitly studied in the following.

Passive Scanning

The passive scanning mode specifies that a station does not start any transmission in order to find other access points in the vicinity. It merely switches through the different channels and listens for Beacon frames indicating the presence of an access point. If such a Beacon frame is received, the station stores the SNR, the channel the access point is operating in, and other related information. Once the scanning procedure is finished, the station chooses the access point with the highest SNR and initiates the handover by issuing a Reassociation Request.

By default, the interarrival time of a Beacon frame at an access point is set to 100 ms. If the station performs the normal passive scanning operation, all channels are scanned individually. Considering the 13 allowed channels in most of Europe, the scanning mechanism at least takes 1300 ms. In fast passive scanning mode, where only the non-overlapping channels are scanned, the whole operation needs no less than 300 ms. Therefore, even the fast passive scanning operation is too slow for most QoS demanding real-time applications. However, the scanning time can be further reduced, if the interarrival time of the Beacon frames is chosen to be shorter than 100 ms. This is a valid approach, since the IEEE 802.11 standards do not demand such a setting.

In Figure 5.62 the results are shown for the varied Beacon Interarrival Times. The dashed line shows the total handover time in milliseconds. It is directly proportional to the Beacon Interarrival Time. In case of a Beacon Interarrival Time of 5 ms, the handover can be finished after about 20 ms while a 100 ms Beacon Interarrival Time leads to a total handover time of about 305 ms.

However, choosing a smaller Interarrival Time for the Beacon frames has a drawback as well. This can be shown by evaluating the maximum throughput that can be achieved in the different scenarios. In our case, the station acts as a saturated UDP source, such that it utilizes the whole bandwidth that remains. This maximum achievable throughput is shown as the solid line in Figure 5.62. With 100 ms Interarrival time, more than 5.5 Mbps can be reached. Setting it to 50 ms leads to a reduction of just about 50 Kbps, which is quite acceptable considering the fact that the time for a handover is roughly cut in half.

However, the maximum achievable throughput drastically decreases once the Beacon Interarrival Times are set to a value lower than 20

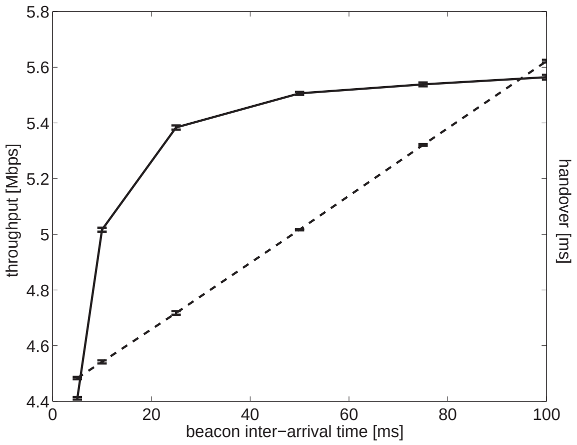

Figure 5.62: *Fast Passive Scanning with varied Beacon Interarrival Times*

ms. Therefore, values of around 50 ms and fast passive scanning are a good tradeoff between maximum achievable throughput and handover performance. In this case the handover can be performed within 150 ms, which is sufficient for QoS demanding real-time applications.

Active Scanning

In case of Active Scanning, the handover performance does not depend on the Beacon Interarrival Time, but on the two timers MinChannelTime and MaxChannelTime. As explained earlier, the station transmits a Probe Request frame on the scanned channel and waits for Probe Response frames for a period of at least MinChannelTime. If no activity is detected, the station moves on to the next channel. If, on the other hand, a Probe Response frame is received, the station keeps on listening on the channel for a period of MaxChannelTime. This is due to the fact that multiple access points might be set to the same or an overlapping channel, such that multiple Probe Response frames can be received on a single channel.

Therefore, the MaxChannelTime has a direct influence on the duration of the active scanning procedure. The goal is to minimize it as much as possible. However, Probe Response frames are send using the standard DCF operation. The more answers a station receives on a specific channel, the longer the MaxChannelTime should be chosen. This fact can be seen in Figure 5.63. It shows the maximum delay in case of a varying number of Probe Response frames received by the station on a single channel.

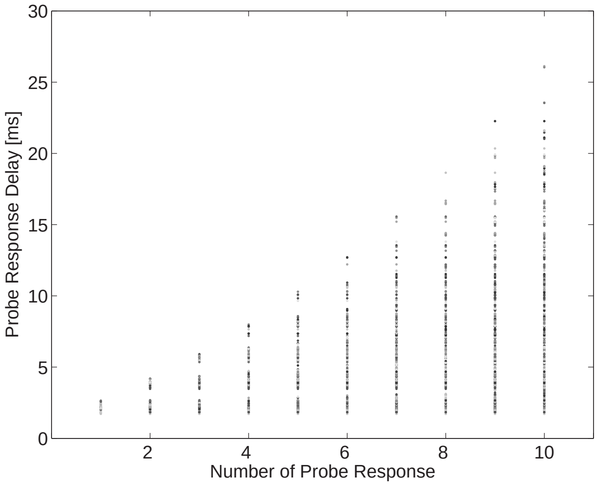

Figure 5.63: *Probe Response Delays in Active Scanning Mode*

Number of responses	Maximum Delay	Access Point density	MaxChannelTime
1	2.6 ms	low	7 ms
2	4.2 ms		
3	5.9 ms		
4	8.0 ms	medium	17 ms
5	10.3 ms		
6	12.7 ms		
7	15.6 ms		
8	18.6 ms	high	27 ms
9	22.3 ms		
10	26.1 ms		

Table 5.14: *Active Scanning MaxChannelTime setting based on access point density*

In the case of just one Probe Response frame, the delay only varies between about two and three milliseconds. On the other hand, if ten access points answer the station's call, up to 27 milliseconds are necessary to receive all the responses. It is necessary for a station to receive as many of the answers as possible, since these answers directly influence the handover decision.

One way to optimize the active scanning procedure is to set the MaxChannelTime according to the access point density, as shown in Table 5.14. If there are many access points surrounding the station, the MaxChannelTime should be chosen high, while a low access point density allows shorter time periods.

Such a setting allows the station to still receive all the Probe Response frames, while long periods of inactivity due to too large MaxChannelTime scanning periods are avoided. The MinChannelTime on the other hand is the same for all scenarios. The whole scanning time, thus, solely depends on the access point density setting. When normal active scanning is performed, the scanning procedure in the 13 channels of most European countries sums up to a total of 91 ms in low density cases and 351 ms in high density areas. When fast active scanning is performed, the scanning time ranges between 21 ms and 81 ms, which is acceptable for QoS demanding applications.

Neighborhood Detection

In the case of scanning with neighborhood detection, a station does not have to scan all the available channels. As explained earlier, the potential destinations of a handover are specified within the Beacon and Probe Response frames of each access point. A 20 Byte block is added to these frames for each access point that is announced, which leads to larger packet transmission times.

In addition all of these access points have to be scanned by the station in order to receive SNR information. Figure 5.64 shows the handover delay for a varying number of announced access points. It ranges from 4 ms to about 23 ms on average depending on the number of access points which is varied from two to twelve. No more than twelve access points can be announced within a single Beacon or Probe Response frame.

As expected, scanning with neighborhood detection turns out to be the fastest handover mechanism. On the downside, it is the most complex approach. And there are still some issues that need to be resolved as was discussed earlier.

Comparison of the Handover Mechanisms

The last sections evaluated the handover mechanisms using the different scanning approaches. Large differences were found and some possibilities to optimize the scanning mechanisms were discussed. Table 5.15 summarizes the results. Each of the handover mechanisms was simulated 50

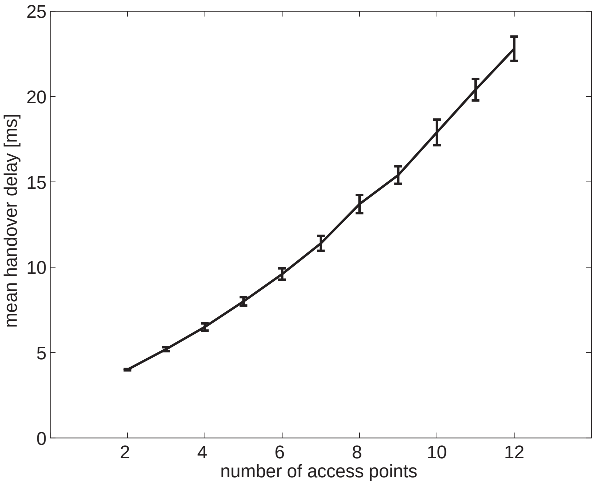

Figure 5.64: *Handover delays with Neighborhood Detection*

times, and the average of the results was calculated.

The results show that Neighborhood Scanning performs best. It is appropriate for all the desired QoS demanding applications considered here. Handover delays of approximately 21 ms on average can hardly be noticed by the user. The same holds for fast active scanning. It is about 50 percent slower, but an average handover delay of 31 ms is still quite appropriate. Voice and video transmissions require one-way delays of less than 150 ms according to the ITU-T Study Group 12. Therefore, the normal active scanning as well as the fast passive scanning with handover delays of 140 ms and 150 ms are rather large. Considering additional delay due to the wired backbone network or the coding delays can lead to bad quality of the transmission. However, fast passive scanning is inappropriate, if the Beacon Interarrival Time is left at the standard

Scanning Mechanism	Scanning	Reassociation	Full Handover
Passive	650 ms (1300 ms)	2.649 ms	652.549 ms (1302.649 ms)
Fast Passive	150 ms (300 ms)	2.631 ms	152.631 ms (302.631 ms)
Active	140.393 ms	2.616 ms	143.009 ms
Fast Active	32.335 ms	2.593 ms	34.928 ms
Neighborhood	21.369 ms	2.315 ms	23.684 ms

Table 5.15: *Active Scanning MaxChannelTime setting based on access point density*

value of 100 ms. Finally, normal passive scanning is by far too slow. Even with the an adapted Beacon Interarrival Time the handover delay is no less than 650 ms.

5.5.2 Voice traffic

The last section showed that the performance of a handover in Wireless LAN mainly depends on the time spent for scanning. Two of the five scanning mechanisms turned out to be very fast. Full handover delays of just 23 ms and 35 ms were possible. However, no background traffic was considered. In this section we explicitly study the handover performance in case of voice clients. The simulation scenario is shown in Figure 5.65. Several stations are located in each of the different cells while one station roams between the three attached access points. The access points are 70 meters apart from one another. They use different non-overlapping channels, i.e. access point AP1 uses channel one, access point AP2 transmits on channel six, and access point AP3 is set to channel eleven.

As pointed out in Section 5.3.2, the best voice codec in Wireless LAN environments is the G.723.1 standard. It allows up to 18 simultaneous

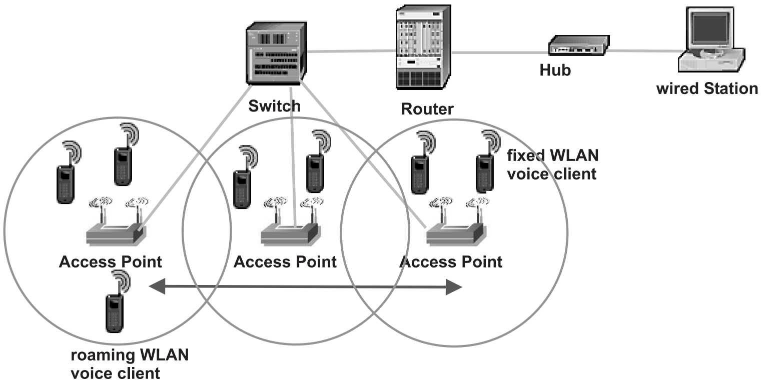

Figure 5.65: *Simulation scenario with multiple voice clients*

voice clients within a single cell. The voice clients in the following simulations are connected to a voice client in the backbone network.

The first scenario is that there is a single station moving back and forth. No other stations are considered, i.e. no background traffic occurs. The results are shown in Table 5.16. Passive scanning was not considered, since handover delays of 650 ms and more are simply too far from being suitable for voice data. The table summarizes the number of lost packets during the handover and the number of forwarded packets for the remaining four scanning mechanisms.

When neighborhood scanning or fast active scanning is used, only one packet gets lost. The handover is performed within such a small period of time, that no packets need to be forwarded by the old access point. On the side of the station only one single packet is delayed for an average of 4.6 ms with neighborhood scanning and about 26.7 ms on average when fast active scanning is performed.

Normal active scanning leads to seven lost packets on average. Voice

Scanning Mechanism	Average Number of		End-to-End Delay		Average Interruption in ms
	Lost Packets	Forwarded Packets	max.	avg.	
Neighborhood	1	0			30
Fast Active	1	0			30
Active	7	1.02	121.3	117.3	210
Fast Passive (50)	10.1	5.8	157.1	122.8	303.6
Fast Passive (100)	19.9	12.0	307.7	232.1	596.4

Table 5.16: *Forwarded packets statistics*

packets are sent every 30 ms, such that the overall interruption sums up to be about 210 ms, which still conforms to the ITU-T requirements [Cov01]. On average one single packet is forwarded with an average delay of 117 ms.

Two different scenarios were studied for fast passive scanning. As was pointed out in the last section, the fast passive scanning mechanism can be optimized by setting the Beacon Interarrival Time to 50 ms instead of the 100 ms default setting. The results for both configurations are shown. For the 50 ms Beacon Interarrival Time, about 10 packets get lost, leading to a service interruption of about 300 ms. On average about 6 packets are forwarded after the successful handover, with a delay of 123 ms. Such a performance is still acceptable according to the ITU-T recommendations. If the 100 ms Beacon Interarrival Time is used, on the other hand, about 20 packets are dropped and 12 packets are forwarded. The mean interruption of the service with almost 600 ms is, however, too large for an acceptable voice quality.

Next we want to study the effect of background voice traffic on the handover performance of a single voice client. Therefore, a varying number of fixed stations acting as voice clients are simulated while one single station is roaming between the access points. Again, the normal passive scanning procedure is not considered, because it is by far too slow.

The results are shown in Figure 5.66. It compares the handover delays for the different scanning mechanisms. The number of background voice clients is varied from zero to 17.

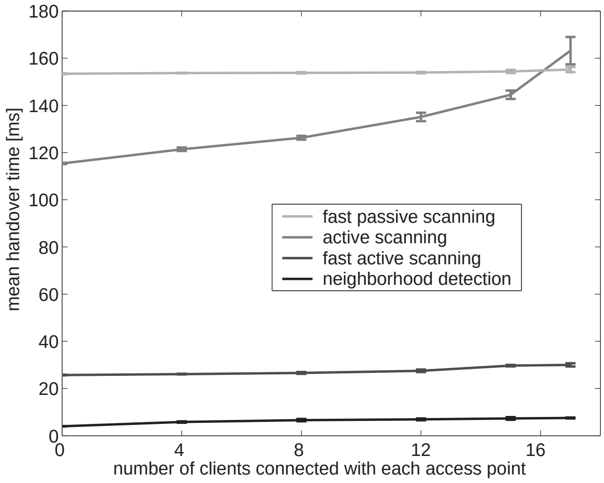

Figure 5.66: *Handover delay with background voice traffic*

Again, neighborhood detection is the fastest scanning mechanism. Handovers are performed within 4 ms to 7.5 ms and the number of background voice clients has just a minor effect. The same holds for fast active scanning. The number of background users just has a minor effect on the performance. Handovers are finished after 26 ms to 30 ms. All these handovers are very fast and voice users do not notice any interruption of the service.

In contrast to these two scanning mechanisms, the active scanning procedure is by far more influenced by the number of background voice traffic. Therefore the performance of the handover decreases with the number of concurring voice clients, such that for high values, the performance of an active scanning handover even gets worse than if fast passive mode with 50 ms Beacon Interarrival Time is used. Handover delays range from about 117 ms to approximately 167 ms. However, fast active scanning as well as passive scanning can still provide acceptable results. Minor distortions of the voice stream can occur, but the voice quality still reaches acceptable levels.

Summarizing the results for voice application, we can conclude that several of the handover mechanisms are appropriate even if there are multiple voice clients simultaneously active in the network. As expected, neighborhood detection works best, but the fast active scanning almost reaches the same quality. The remaining procedures, active scanning and fast passive scanning, still perform well enough. However, some minor disturbances occur.

5.5.3 Video traffic

Handover mechanisms in the case of voice traffic with its constant bit rates, were found to work quite well in the last section. Now, we turn to video traffic with variable bit rates. The simulation scenario is shown in Figure 5.67. Again a number of video clients is placed within the different cells. All of these stations do not move. One additional station roams through the Wireless LAN coverage area and performs frequent handovers. The goal is to evaluate the effect of variable bit rate real-time applications on the handover performance.

As explained in Chapter 3.2.4, CIF and QCIF videos are transmitted from the video clients in the backbone network to the WLAN stations and back. This simulates the behavior of interactive video conferences.

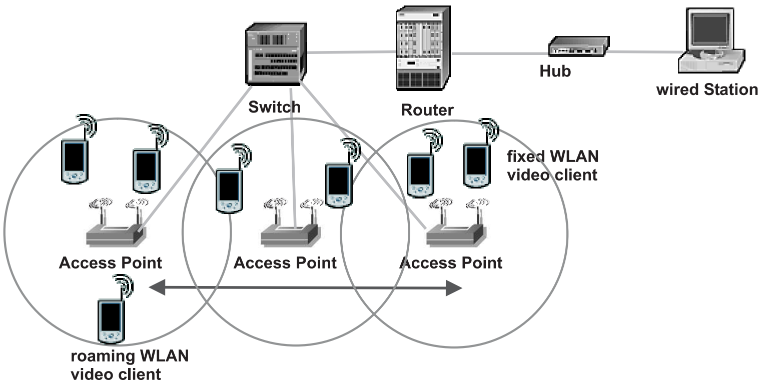

Figure 5.67: *Video traffic simulation scenario*

In order to evaluate the neighborhood scanning mechanism, a maximum of twelve access points is placed in the network. Neighborhood scanning allows the indication of up to 12 access points within the Beacon and Reassociation Request frames of the access points. All of these access points have to be scanned by the station, such that neighborhood scanning performs worst if the maximum number of access points is used. This allows us to simulate the worst-case scenario for neighborhood scanning.

All other scanning mechanisms that are studied for the video application are simulated using the three access points. As for the case of voice traffic, the access points use the non-overlapping channels one, six, and eleven. The clients use video A from Table 3.7. It utilizes the largest average throughput and has the largest variance. All stations start their video conference at exactly the same time. This defines the worst-case scenario.

Figure 5.68 summarizes the results. It shows the handover delay for the video simulation using the different scanning mechanisms. The number

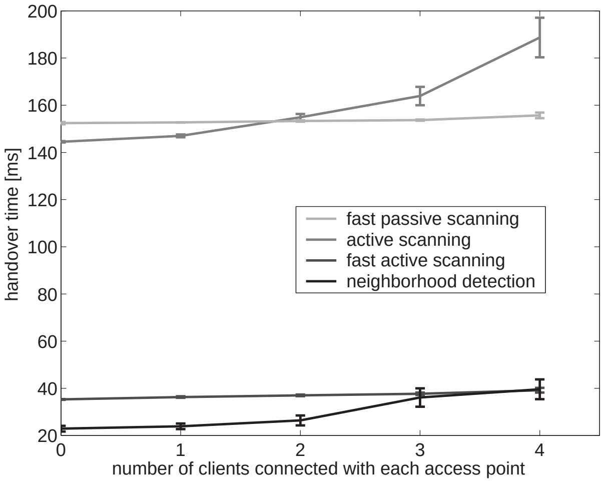

Figure 5.68: *Handover delay with background video traffic*

of background video clients is varied from zero to four.

If there is no other than the roaming station in the network, the handover delay ranges from 22.9 ms for neighborhood scanning to 152.4 ms for fast passive scanning. Again, neighborhood scanning performs best. However, if the maximum of twelve access points is indicated to the station, the handover for neighborhood scanning ranges from 5.1 ms to 75.4 ms. Here, a large variance of the handover performance becomes obvious. This is due to the fact that the station scans the access points until it finds a good candidate, i.e. if the first Probe Request already results in an answer of an access point, the station quits the scanning and starts the handover instantly, leading to a handover delay of just 5.1 ms. However, if the station has to scan all twelve access points until it receives an acceptable answer, the handover delay reaches up to 75 ms.

Scanning Mechanism	Average Number of dropped packets without forwarding		Number of dropped packets with forwarding	Average Service Disruption in ms
	CIF	QCIF	CIF and QCIF	
Neighborhood	1.00	0.73	0-1	40
Fast Active	1.45	1.06	0-1	58
Active	5.92	4.34	0-4	236.8
Fast Passive (50)	6.25	4.58	0-4	250.0

Table 5.17: *Forwarded packets statistics*

If more stations are added to the simulation scenario in the form of fixed video clients, all handover delays increase. In the case of fast passive scanning and fast active scanning, however, this increase turns out to be rather small with just about 4 ms additional delay on average. Fast active scanning performs quite well with less than 40 ms. Fast passive scanning always leads to delays of about 150 ms to 160 ms. The mean handover delay with neighborhood scanning increases from about 23 ms to about 39.6 ms, which is just above the delay for fast active scanning. Normal active scanning experiences the largest degradation. While it lies just above 140 ms if no background traffic is simulated, it increases to almost 190 ms on average.

Table 5.17 summarizes the average number of dropped and forwarded packets in the case of background video traffic. Video A in CIF format and video C in QCIF format were simulated.

Neighborhood scanning and fast active scanning, thus, lead to service interruptions of 40 ms and 58 ms, respectively. Just about one single packet gets lost during the handover. In case of forwarding, this number further decreases. Therefore, these two scanning schemes perform equally well.

Active scanning without forwarding on the other hand leads to about six dropped packets in case of the CIF video, and around 4 dropped packets in the QCIF case. If the forwarding of buffered packets from the old to the new access point is enabled, anything between zero and four dropped packets can be experienced. The service disruption with active scanning reaches about 240 ms, which is quite acceptable according to the ITU-T recommendations.

Similar results can be found for fast passive scanning with a Beacon Interarrival Time of 50 ms. The total amount of time that the video service is disrupted sums up to 250 ms on average. Considering the 400 ms limit specified by the ITU-T, this mechanism is still quite acceptable.

These simulation results allow to draw the conclusion that even in the case of video traffic, several scanning techniques and, therefore, several handover mechanisms are fast enough to perform handovers without disrupting the service. In the next section a traffic mix is studied. The goal is to find the handover mechanisms that are fast enough, no matter what kind of clients are active. The only prerequisite is that the system is not overloaded.

5.5.4 Traffic mix

The last sections studied the effect of voice and video traffic, while there is traffic of the same type in the background, i.e. voice and video traffic respectively. It turned out that even in highly loaded systems, there are handover mechanisms that perform well.

In this section, a traffic mix is studied. The simulation scenario is similar to the voice or video simulation scenario. It is shown in Figure 5.69. Three access points are considered. They operate on the non-overlapping channels one, six, and eleven. The access points are 70 meters apart, a handover is performed once the roaming station moves more than 50 meters away from the access point it is associated with.

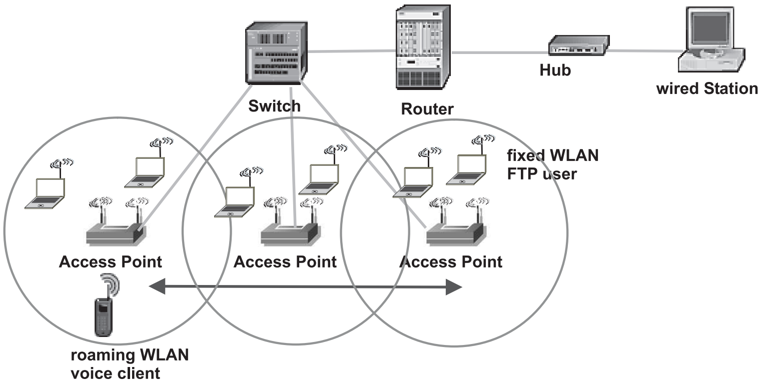

Figure 5.69: *Traffic mix simulation scenario*

The background traffic is generated by a number of fixed stations positioned at each access point. In order to consume all the available bandwidth, these fixed stations perform FTP downloads in this section. The FTP server is positioned in the wired backbone network. FTP downloads can be thought of as a kind of worst-case HTTP traffic. Web users have short periods of time when they perform downloads of objects. In these time periods, the whole available bandwidth is used up. An FTP user on the other hand, consumes the whole bandwidth all the time.

In order to study handover performance, a roaming station is considered that performs a voice transmission with a voice client in the backbone network. This roaming station moves back and forth within the access points as indicated in the figure. The different scanning mechanisms are analyzed. The normal passive scanning mechanism is not sufficient for any type of real-time traffic, since it usually takes about 650 ms. Therefore, as in the last section, normal passive scanning is not studied.

The cumulative handover delays are displayed in Figure 5.70. It shows that when fast passive scanning is used, by far more than 90 percent of the handovers are finished after 155 ms. However, there are hardly any faster handovers. The fast passive scanning usually takes around 150 ms and 160 ms and the variance is very low.

Normal active scanning, on the other hand, shows a quite different behavior. The variance is by far larger. The handover delay ranges from 125 ms to about 225 ms. In about 40 percent of the cases, the normal active scanning is faster than the fast passive scanning, but in 60 percent of the cases it is slower. Therefore, fast passive scanning shows a nicer behavior than the normal active scanning.

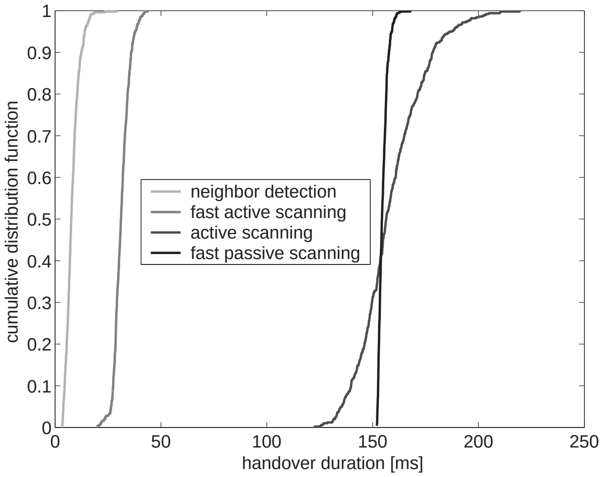

Figure 5.70: *Handover delay with traffic mix*

Nevertheless, normal active scanning as well as fast passive scanning are by far slower than the remaining two scanning mechanisms. Using fast active scanning, the handover delay ranges from about 20 ms to approximately 45 ms, which is roughly three times faster than the other two procedures. Fast passive scanning has a small variance, which is a desired behavior.

As seen before, the neighborhood scanning mechanism is again the fastest scanning mechanism. In the traffic mix scenario, the fastest handovers are finished after about 5 ms, while most of the handovers are finished after no more than 20 ms. In rare cases, the full handover procedure might take about 40 ms. The variance is still very small.

The simulation scenarios considered here, only used three access points. However, we have seen in the last section already, that the neighborhood scanning mechanism is susceptible to the (announced) number of access points. The active scanning mechanism, on the other hand, is not influenced if we add more access points to the scenario. In addition, the neighborhood scanning mechanism is by far the most complex procedure and is not defined in the Wireless LAN standards.

Therefore, the fast active scanning mechanism should definitely be proposed as the default way of performing the scanning procedure in QoS enabled Wireless LAN environments as they appear in the context of 4G networks.

5.6 Conclusions

The goal of this chapter was to evaluate the QoS capabilities of the Wireless LAN MAC protocol and its extensions in future 4G environments. As a practical example, a large office building can be considered. If such a large-scale environment has to be implemented using Wireless LAN, several issues arise that need to be taken care of.

First of all, the 2.4 GHz frequency band is considered the better alternative than the 5 GHz band, since coverage in the lower frequencies is far easier to reach. However, Wireless LAN in the 2.4 GHz frequency band defines up to 14 channels, but only three are non-overlapping. Overlaps in terms of frequency cause interference and limit the throughput. In large-scale implementations, such overlaps can not be avoided completely. Therefore, the impact of overlapping cells was studied extensively.

Several different scenarios had to be taken into account. In all of these scenarios, the Wireless LAN MAC protocol has to support QoS demanding applications. Our studies showed that the interference caused by overlapping cells can cause great problems. In certain situations, even best-effort users can block each other completely due to the high probability of collisions. However, it was shown that the QoS enabling Hybrid Coordination Function can be configured in a way to support service differentiation to solve these problems.

It was shown that the proposed default parameters for HCF are not sufficient. They do not allow the necessary level of prioritization. Different priority classes were introduced that support QoS in all the considered scenarios. An algorithm was proposed that assigns priority classes automatically as needed.

These studies showed that voice, video, and best-effort traffic can be performed simultaneously in overlapping and co-located cells, while still supporting the necessary QoS levels. However, all the stations in these scenarios were fixed and had no need to perform a handover to a different access point.

Therefore, handover scenarios were considered in the remainder of this chapter. It turned out that the most important task during a handover is the scanning procedure. The Wireless LAN standard proposes a number of different scanning mechanisms. All of these scanning mechanisms were studied in detail. In addition, a promising scanning approach discussed in the literature, was also considered.

Various scenarios were studied. Again the goal was to evaluate the capabilities of the mechanisms to support QoS demanding applications. Several of these scanning mechanisms were proven to provide the necessary functionality. However, in large office buildings with a high number of access points, neighborhood scanning should not be implemented, because fast active scanning with a total handover time between 35.3 ms and 39.2 ms is sufficient and performs even better in highly loaded networks. Therefore, the stations should be configured to perform fast active scanning as the default scanning mechanism and if no access point is found, the station shall switch to fast passive scanning.

However, QoS on any MAC layer can only be provided as long as the traffic load in the system is kept at a certain level. In our studies we found out that using the G723.1 codec for voice traffic, up to 18 simultaneous voice clients can be supported. However, if there are 19 or more stations in the cell trying to perform voice transmissions, the system gets overloaded and the QoS can not be supported adequately. Therefore, an admission control mechanism is necessary to make sure that the load is kept at an acceptable level. It is left for future studies to solve this issue.

The main conclusion of this chapter is that it is definitely possible for the Wireless LAN Hybrid Coordination Function (HCF) to support QoS even in large-scale implementations. Handovers can be realized fast enough to support voice and video applications by implementing the fast active scanning mechanism which is defined in the Wireless LAN standard. Therefore, Wireless LAN has the required capabilities to be integrated into future 4G networks on the Logical Link Control layer.

6 IP - Internet Protocol and Mobility

> I do not fear computers. I fear the lack of them. Isaac Asimov (1920 - 1992)

The last chapter dealt with Quality of Service support of the Wireless LAN Medium Access Control protocol when large-scale scenarios are considered and handovers on ISO/OSI layer two are taken into account. However, the size of a network, where only the capabilities of the MAC protocol are needed for QoS support, is rather restricted. Such local IP subnets are usually restricted to a building floor or at maximum the whole building itself.

Human mobility, however, by far extends these geographical regions, especially when mobile devices become smaller and almost as powerful as desktop computers. Therefore, the QoS handover support has to be lifted to the IP layer as well.

IP mobility has been the topic of many researchers for several years. The basic approach by Charles E. Perkins even dates back to the year 1996 [Per96]. It could not support any type of QoS and it exhibits some other major problems, but it still forms the basis of today's proposals.

In this chapter, we introduce the most important approaches and discuss their advantages and drawbacks considering a Wireless LAN network with QoS support. One such scenario is shown in Figure 6.1. It consists of two IP subnets that are both connected to the Internet.

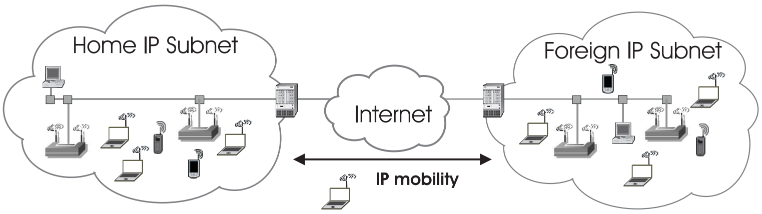

Figure 6.1: *IP handover scenario*

Our major interest lies in the fact that once a wireless device roams from one of these IP subnets to the other, it is not sufficient to simply perform a handover on MAC layer, where only the associated access point of a station is changed. In the case of a movement between different IP subnets, the routing mechanisms of the Internet as well as the Intranet require a change of the local IP address of the mobile device. However, a seamless change of the local IP address is not supported by most of the currently implemented applications. An update of all these applications is not feasible, such that other approaches become necessary.

The second possible solution is to propagate host-specific routes for each mobile node. However, considering the explosive growth of the number of mobile devices, such a mechanism is impractical as well. Therefore, new proposals are necessary. In addition to the mere change of the IP address, an integrated 4G network requires the IP handover to be performed within stringent time limitations. In the following the most important proposals for IP mobility are presented and their capabilities to solve the IP handover problem in future 4G networks are studied.

These IP handover mechanisms are greatly discussed for the upcoming IPv6 protocol. It will have an integrated support for IP mobility. On the other hand, IPv4 does not have the flexibility to support the protocol itself. However, the basic mechanisms of IPv4 and IPv6 mobility support are rather similar, such that it is sufficient for our purposes to solely consider the IPv4 proposals.

6.1 Mobile IP

The first proposal for IP Mobility Support was released in 1996 as RFC 2002 [Per96]. In the meantime, several changes and extensions were added, such that the current version of the document is RFC 3344: *IP Mobility support for IPv4* [Per02].

It defines a way to always identify a node by its home IP address, regardless of its current location. If the node is away from home, it is associated with a care-of address, which provides the necessary information to route the packets through a tunnel to the mobile node.

6.1.1 Mobile IP architecture and procedures

The architecture of the Mobile IP protocol is show in Figure 6.2. It consists of three different components. The *Mobile Node* is defined as a host or router that changes its point of attachment from one IP subnet to another. During a handover, it may continue to communicate with other nodes in the Internet, the Correspondent Nodes (CN).

The *Home Agent* is a special router in the Mobile Node's home network. It tunnels the datagrams for delivery to the Mobile Node, when it is located in a foreign network. The *Foreign Agent* is a router in the visited network. It provides the routing services to the Mobile Node. This means, it detunnels the datagrams from the Home Agent and delivers them locally to the Mobile Node.

Therefore, the communication of the Mobile Node to the CN located in the Internet is performed as indicated by the dashed arrows in Figure 6.2. Datagrams initiated at the Mobile Node are routed to the CN (1). Routing is performed according to the destination address. The routes,

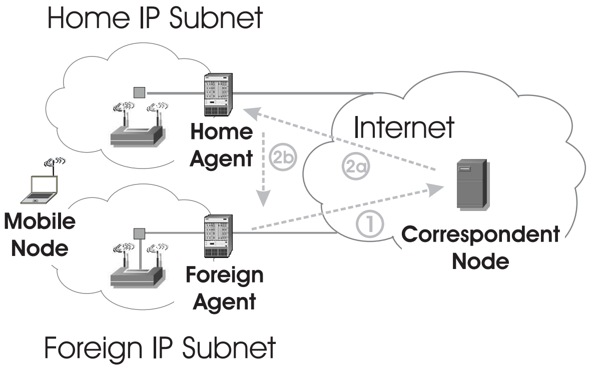

Figure 6.2: *Mobile IP architecture and packet flow*

thus, do not consider that the Mobile Node's IP source address is not within their IP subnet.

The CN simply answers to the requests by transmitting the datagrams back to the source IP address of the received packets. Therefore, the routing mechanisms in the Internet make sure that the packets arrive at the Mobile Node's home IP subnet (2a).

The Home Agent stores a list of Mobile Nodes that are currently away from home. The entries of this list consist of the home IP address and the corresponding care-of address of the Mobile Node and are called *Mobility Bindings*. They allow the Home Agent to decide whether the datagrams simply should be forwarded to the home IP subnet or if they need to be tunneled to a Foreign Agent (2b).

The list of the Mobility Bindings in the Home Agent is updated by the Registration service implemented in the Mobile Node. As soon as it is away from home, it registers its care-of address with its home agent. This can be done directly or through the Foreign Agent, depending on the method of attachment as described later.

In order for a Mobile Node to find a Foreign Agent in a visited network, two different methods are implemented. All Home and Foreign Agents may advertise their availability by issuing Agent Advertisement messages. In the initial version of the Mobile IP protocol, the interarrival time of these messages was set to at least one second. Therefore, upon a successful association to a foreign Wireless LAN network, the average delay until a reception of an Agent Advertisement message was 500 msec, which is too long for video and voice services. Therefore, the interarrival time of the Agent Advertisement messages can be set to smaller values as well in the newer version of the Mobile IP protocol [Per02].

In the second case, the Mobile Node plays a more active role. After it enters a new network, it issues an Agent Solicitation message. The Home and Foreign Agents answer the request by sending an Agent Advertisement. This way, the delay until the necessary information of the connected network is found, can be greatly reduced. The two mechanisms can be compared to the active and passive scanning mechanisms of the Wireless LAN MAC protocol.

Once the Mobile Node retrieved the necessary Foreign Agent information, it can start the care-of address acquisition process. Two different methods are proposed. In the first mode, the Foreign Agent propagates its own IP address as the care-of address. The Foreign Agent itself becomes the endpoint of the tunnel from the Home Agent. This mode is referred to as the *foreign agent care-of address*. This mode is preferred, since many Mobile Nodes can share the same care-of address and therefore put less demand on the already limited IPv4 address space.

The second mode is called *co-located care-of address*. Here, the Mobile Node acquires a local IP address through some external means, such as the Dynamic Host Configuration Protocol (DHCP). It then associates its own network interface with this local IP address. In this mode, the Mobile Node is the endpoint of the tunnel to the Home Agent and has to take care of decapsulation of the datagrams. In this mode, no Foreign Agent is necessary, but more global routable IP addresses have to be accessible. Consider that the co-located care-of address has to be routable from the Home Agent. Therefore, using this mode of operation within a IP subnet that performs Network Address Translation (NAT) to the outside, is not an option!

6.1.2 Mobile IP performance issues

The basic IP mobility support was designed to allow Mobile Nodes to roam to foreign IP subnets. However, the performance of these IP handovers was of minor importance. The mere goal was to provide the necessary functionality. Therefore, several issues arise that have a negative impact on the handover performance. These issues are discussed in the following.

Triangle Routing

As indicated by the architectural overview of the Mobile IP mechanism in Figure 6.2, the data traffic of a Mobile Node located at a foreign network and communicating with a Correspondent Node somewhere in the Internet exhibits *triangle routing.* Data packets initiated by the Mobile Node are directly transmitted to the Correspondent Node, since the routing protocols base their decisions on the destination address.

The Correspondent Node answers the requests by issuing packets to the source address, which is the home IP address of the Mobile Node. Therefore, the packets are routed to the home IP subnet. From there, the packets are delivered to the Mobile Node using tunneling. As explained earlier, either the Foreign Agent or the Mobile Node itself is the endpoint of the tunnel. In either case, the packets need to be tunneled to the foreign IP subnet.

In terms of performance, the triangle routing definitely leads to an increase of the network delay. In addition to the normal delay, the packets always have to traverse the tunnel from the home to the foreign network.

Micro-Mobility versus Macro-Mobility

Once a Mobile Node roams to a foreign subnet, it initiates a handover. First the local registration process is initiated. As explained above, there are two different methods that allow the Mobile Node to associate with the foreign IP subnet. In case of a foreign agent care-of address it has to learn the necessary infrastructure information about the visited network, while in the case of co-located care-of address, the Mobile Node needs to retrieve a global IP address from the foreign IP subnet.

After the local registration in the foreign network is completed, the Mobile Node needs to perform a *Binding Update*, which informs its Home Agent of its current care-of address. In turn, the Home Agent can now start to setup the tunnel for the Mobile Node. This completes the Mobile IP handover and allows the Mobile Node to continue its communication with the Correspondent Node.

Such a handover process is frequently referred to as *Macro-Mobility*. It defines that the Mobile Node changes its point of attachment in a way that a binding update at its Home Agent becomes necessary. Three cases can be distinguished that demand for Macro-Mobility. Either the Mobile Node moves away from its home IP subnet to a foreign subnet, or it roams from one foreign IP subnet to another, or it simply returns home after visiting a different network. In all of these cases, the Home Agent needs to update its binding list and the related tunnel.

During the Macro-Mobility procedures, the Mobile Node experiences a certain amount of time where it can not communicate with the Correspondent Node. This handover delay depends on the time for the local registration and the tunnel establishment. QoS demanding applications such as voice conferencing prefer one-way delays of less than 150 msec. At most a one-way delay of 400 msec is acceptable. Therefore, Macro-Mobility handovers are critical if a QoS level has to be provided.

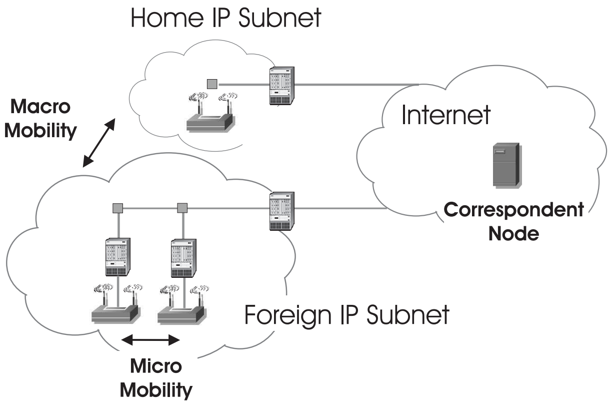

Figure 6.3: *Macro-Mobility versus Micro-Mobility*

However, in large IP subnets a different type of mobility can be considered. Since large-scale IP subnets are usually separated into logical subgroups, there are often several Foreign Agents located in a single IP subnet. In such cases, the utilization of Mobile IP leads to a higher frequency of Macro-Mobility handovers that are actually not necessary. Optimizations of the standard mechanisms are possible. These are not considered in the IP Mobility Support RFCs, but it is mentioned that the pure Mobile IP functionality is not sufficient for Micro-Mobility support. In the next section such Micro-Mobility enhancements are explained. They help to reduce the organizational overhead of the protocol and therefore lead to a reduced delay.

The difference between Macro-Mobility and Micro-Mobility can be seen in Figure 6.3. It shows that a movement between two subnets requires a Macro-Mobility handover while a movement within a single subnet can be covered by Micro-Mobility support.

Agent Advertisements and Solicitation Messages

The local registration process of a Mobile Node roaming to a new network itself can be distinguished in terms of handover performance. As described above there are two different ways a Mobile Node can learn about the local infrastructure. Either it scans the network traffic for regularly transmitted Agent Advertisements, or it actively broadcasts a Soliciation Message to get informed of the Home or Foreign Agent.

In the case of Agent Advertisements, the Home or Foreign Agents, frequently broadcast data packets that inform the stations in the local IP network of their presence. In addition, other relevant information is transmitted, such as the lifetime or the router address. The original RFC 2002 specified a advertisement frequency of no less than one second. However, this leads to an average delay of 500 msec for Mobile Nodes entering the network. This is by far too large for many real-time applications. Therefore, the newest RFC 3344 allows higher frequencies. Nevertheless, higher advertisement frequencies have the drawback of increasing the overhead of the protocol, such that a good trade-off between advertisement frequency and protocol overhead has to be found.

Solicitation Messages, on the other hand, allow a Mobile Node to perform an active search for Mobility Agents in the currently associated IP network. The RFC 3344 defines that such Solicitation Messages should only be sent in the absence of Agent Advertisements. After successful reception at a Mobility Agent, it responds by transmitting an Agent Advertisement, which in turn informs the Mobile Node of the necessary information.

In terms of performance, the Solicitation Messages can be used to reduce the initial delay of IP handovers. The term "SHOULD" in the RFC allows to do so, if *understood and carefully weighed*. Therefore, it seems appropriate to perform such active scanning at least for QoS demanding Mobile Nodes, in order to keep the initial delay as low as possible.

6.2 Mobile IP Enhancements

In this section, several important enhancements to the basic Mobile IP proposal are presented. Due to the vast number of proposals that are known from the literature, it is impossible to even name all of them. Therefore, just a small number of extensions were chosen.

6.2.1 Route Optimization in Mobile IP

Charles E. Perkins and David B. Johnson proposed a route optimization scheme for Mobile IPv4 [PJ02]. The first version dates back to the year 1994, while the last version was issued in 2002. It is still considered as work in progress and has not been released as an RFC by the IETF. However, first implementations exist and several performance studies were published.

The main idea is to create a way to get around the triangle routing as it happens when the basic Mobile IP protocol is applied. In order to do so, a binding cache is introduced that allows to store the current care-of address of the Mobile Node and to tunnel the data packets directly to this care-of address bypassing the Mobile Node's Home Agent.

A binding cache can be kept by any node. The entries of the binding cache are inserted every time an authenticated *Binding Update* message is received. These messages do effectively change the routing of datagrams, such that the authentication of such messages is important.

If no binding cache is used, the datagrams are send as specified by the basic Mobile IP protocol. The Correspondent Node transmits the packets to the Mobile Node's home IP address. There they are intercepted by the Home Agent and tunneled to the current care-of address. The route optimization extension now specifies, that every time a Home Agent receives datagrams and tunnels them to the foreign location, it also sends a *Binding Update* message to the originating address of the Correspondent Node.

Correspondent Nodes that do not keep a binding cache may simply ignore those messages. However, it may also insert a new entry to its local binding cache. This entry informs the Correspondent Node of the current care-of address of the Mobile Node. It can then be used to directly transmit the data packets to the current point of attachment. In order to do so, the Correspondent Node has to set up a tunnel as the Home Agent does.

Problems arise, if the Mobile Node changes its point of attachment, once the Corresponding Node performs route optimization. The packets are still transmitted to the old care-of address. However, in case of a foreign agent care-of address, the endpoint of the tunnel can deduce that the Correspondent Node uses an out-of-date binding cache entry. It, therefore, informs the Mobile Node's Home Agent of such a node, which in turn transmits a new Binding Update message to the Correspondent Node. Once this Correspondent Node updates its binding cache, the packets can be delivered to the correct care-of address.

In addition, the route optimization extensions also enable smooth handovers, where the old Foreign Agent of a Mobile Node is informed of the new care-of address and can, thus, forward stored datagrams to the new location. In the original approach, the Home Agent changed the care-of address of the Mobile Node that has moved and sets up a new tunnel to the endpoint. However, data packets that were intercepted and sent to the previous care-of address in the time between the handover and the registration update, were lost and had to be taken care of by higher layer protocols.

Route optimization introduces a way to reliably notify the Mobile Node's previous Foreign Agent of its new point of attachment. Therefore, the buffered datagrams can be forwarded to this location. Besides, this mechanism also allows to forward the packets that a Foreign Agent receives due to the out-of-date binding cache entries of Correspondent Nodes. Not only does the Foreign Agent take care of informing the Home Agent to send a Binding Update message to the Correspondent Node, but it also makes sure that the packets do not get lost, but are forwarded to the right location.

It should be mentioned, however, that the route optimization can only be introduced if the foreign agent care-of address mode is utilized. If there is no dedicated Foreign Agent in the foreign IP subnet and the Mobile Node uses the co-located care-of address, out-of-date binding cache entries can not be detected and smooth handover becomes impossible.

6.2.2 Mobile IPv4 Regional Registration

As an optional extension to the Mobile IPv4 protocol, the Mobile IP Working Group defined the *Mobile IPv4 Regional Registration* mechanisms [GJP04]. The main idea is to introduce a layer of hierarchy in a visited domain, such that a fast Micro-Mobility mechanism can be supported. This is necessary since each router that connects IP subnets within this large IP network has to implement the Foreign Agent functionality. Instead of registering with the Home Agent every time a Mobile Node roams to a new location within a large IP subnet, a new entity is introduced, the Gateway Foreign Agent. It takes care of the user mobility within the foreign domain. Therefore, the number of signaling messages to the home network is reduced, which helps to reduce the signaling delay in case of user mobility.

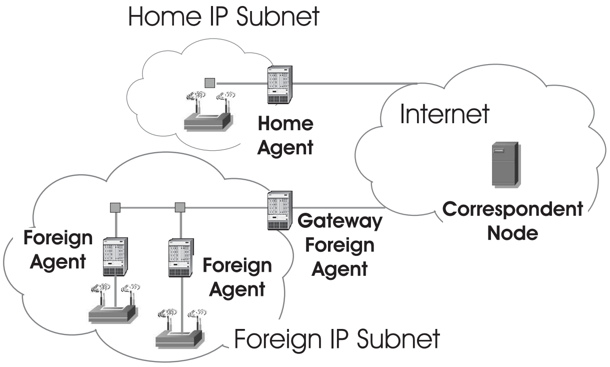

Figure 6.4: *Overview of the Regional Registration support extension for Mobile IPv4*

An overview of the functional components involved in Mobile IPv4 with support for Regional Registration is provided in Figure 6.4. The foreign IP subnet contains a new entity, the Gateway Foreign Agent, which is a Foreign Agent with additional functionality. It forms the basis of a hierarchy of Foreign Agents within the foreign domain and has a publicly routable IP address. Beneath the Gateway Foreign Agent, one or more (Regional) Foreign Agents are used.

If a Mobile Node enters such a foreign network the first time, it has to register its care-of address with its Home Agent as in the basic Mobile IPv4 approach. However, if regional registrations are supported in the foreign network, the address of the Gateway Foreign Agent is registered at the Home Agent. The Gateway Foreign Agent does not change as long as the Mobile Node is located in the foreign domain.

In order to support regional registration, several messages have to be extended. The Registration Requests sent by the (Regional) Foreign Agents to the Gateway Foreign Agent now include a Hierarchical Foreign

Agent extension. This option informs the Gateway Foreign Agent of a regional movement of the Mobile Node, such that it can update its local list for forwarding purposes. The Agent Advertisement messages, telling a Mobile Node about the presence of a Foreign Agent, are extended to indicate the support for regional registration.

The Mobile IPv4 Regional Registration mechanism is downward compatible with the pure Mobile IPv4 proposal. Clients do not have to utilize the extension. However, considering that Micro-Mobility is assumed to happen much more frequently than Macro-Mobility, the reduction of unnecessary handover delays is advisable.

A number of other proposals that help to increase the performance of Micro-Mobility can be found in the literature ([RVS+99], [Val99]). Regional Registration is merely one example.

6.2.3 Low Latency Handoffs in Mobile IPv4

Another interesting Internet draft for the enhancement of Mobile IP performance is the *Low Latency Handoffs in Mobile IPv4* proposal [Mal04]. As discussed earlier, the delay until a Mobile Node receives the necessary information about the local Mobile IP infrastructure depends on the interarrival time of the Agent Advertisement messages. In the original Mobile IPv4 approach, these messages were sent no more often than once per second. The latest version allows for higher frequencies. However, the higher the frequency of these messages, the more load is put on the link. Therefore, a trade-off between the frequency and the additional overhead has to be found.

The *Low Latency Handoff in Mobile IPv4* proposal, on the other hand, introduces a different way to lower the initial handover delay. It defines L2 triggers, i.e. indications initiated on layer two of the ISO/OSI model, to inform the IP layer of a (successful) handover to a new location. Utilizing these triggers, three different methods can be implemented that help to improve handover performance. They are explained in the following.

The methods are *pre-registration*, *post-registration*, and a *combined handover method*. It is worth mentioning that such an approach infringes the clean separation between the different layers of the protocol stack.

When pre-registration is applied, the Mobile Node is assisted by the network in performing an IP handover, before the MAC layer handover is finished. L2 triggers are used by the Mobile Node or the Foreign Agent to initiate particular IP layer events, depending on whether a network-initiated or a mobile-initiated handover occurs.

Basically, pre-registration uses the concept of Proxy Routers, where the Mobile Node is being informed of the new Foreign Agent ahead of the actual handover. The Mobile Node can then start the registration process with the new Foreign Agent, even if a layer two handover has not been performed. All the traffic between the Mobile Node and the new Foreign Agent is routed through the old Foreign Agent as long as the actual MAC layer handover has not been performed. Therefore, the old Foreign Agent acts as a ”proxy router” for the new Foreign Agent. This explains the basic mechanism. However, depending on the type of handover, mobile-initiated or network-initiated, there are slight differences, but the concept stays the same.

Post-registration, on the other hand, is performed after the actual MAC layer handover is finished. L2 triggers are used to set up a bi-directional tunnel between the old Foreign Agent and the new Foreign Agent. Here, the old Foreign agent is referred to as the anchor Foreign Agent. This bi-directional tunnel is used as long as the registration process with the Home Agent is not finished. The anchor Foreign Agent acts as the endpoint of the Home Agent tunnel. Such an approach has two merits. One is that the delay of a handover is clearly decreased. The other advantage is that no packets are dropped at the old Foreign Agent, since they are simply forwarded to the new access point. As soon as the registration with the Home Agent is finished, the new Foreign Agent becomes the tunnel endpoint and normal Mobile IP operation continues.

The third option defined in the *Low Latency Handoff in Mobile IPv4* proposal is a combination of pre-registration and post-registration. These two processes are run in parallel. In this combined solution the pre-registration process is initiated first. If it completes successfully, the proxy mechanism is used prior to the actual layer two handover. However, if the process does not complete before the actual handover is performed, the post-registration is applied. In such scenarios, the post-registration is used as a backup mechanism if the pre-registration fails.

It can be concluded that the utilization of L2 triggers can help to increase the performance of handovers on IP layer. The clear separation of the ISO/OSI layers is undermined, but the proposed solution can easily be integrated into the basic Mobile IP mechanism. The proposal even discusses the integration of the trigger mechanisms in hierarchically structured environments where Gateway Foreign Agents are used to support Micro-Mobility.

6.2.4 Wireless Multiprotocol Label Switching (WMPLS)

Multiprotocol Label Switching (MPLS) defines a switching technology for IP networks that provides the network operator with mechanisms for the engineering of network traffic patterns. The traditional layer three forwarding paradigm is based on independent forwarding decisions at each single hop between a sender and a receiver based on the IP packet header. In MPLS networks, on the other hand, the analysis of the packet header is performed only once when an IP packet first enters an MPLS network. The packet is then assigned to a stream, which is identified by a label. These labels are used as lookup indexes into the label forwarding tables, which in turn stores the forwarding information.

Using MPLS technology, the assignment of the labels to the packets can be done based on various parameters. Therefore, traffic differentiation, QoS level assignment, or predefined routes through the network can be defined. Considering the opportunities of MPLS networks, several publications ([LGT01], [YM01]) can be found that propose mechanisms to exploit the advantages in order to support fast handover procedures with QoS support. Here, just a small set of these proposals is discussed in order to summarize the common ideas.

The basic way of integrating Mobile IP into an MPLS environment was presented in *Integration of Mobile IP and Multi-Protocol Label Switching* [RTFK01]. It was shown, that the switching mechanism can be used to take care of the forwarding of data traffic between the Correspondent Node and the Mobile Node. The IP-in-IP encapsulation is not needed any more. Basically, normal MPLS operation can be used to take care of the necessary functionality. Using MPLS tunnels for the packet forwarding allows to perform traffic engineering on this path, such that the additional delay can be kept at a minimum.

In addition, the authors showed that such an approach is not restricted to a network where the Home Agent and the Foreign Agent are both located within the same MPLS domain. It can be applied to other cases as well. Multiple MPLS domains with edge Label Switching Routers exchanging label information are a possible scenario, as is the case where MPLS domains are connected by an IP cloud in between.

Due to the higher handover rates in micro-cell implementations, such as in Wireless LAN environments, an extension to the basic integration of Mobile IP and MPLS was published in [YM01]. It is called *Hierarchical Mobile MPLS: Supporting Delay Sensitive Applications Over Wireless Internet.*

The basic idea is to improve the performance of local handovers within a single foreign domain. This approach is similar to the Mobile IPv4 Regional Registrations, explained in Section 6.2.2. A hierarchy of Foreign Agents is established within the foreign domain. The top element in this hierarchy is the Foreign Domain Agent (FDA). Whenever a Mobile Node roams within the domain from one Foreign Agent to another, the registration requests are not directly forwarded to its Home Agent with the associated delay, but they are intercepted by the highest level Foreign Agent, the FDA.

IP-in-IP tunneling is not performed, since MPLS tunnels are established, not only between the Home Agent and the FDA, but also within the local hierarchy of Foreign Agents. Therefore, *Hierarchical Mobile MPLS* combines the advantages of Micro-Mobility support with the traffic engineering capabilities of the integrated Mobile IP and MPLS infrastructure. Thus, smoother local handovers can be performed.

6.3 Summary

In this section, some of the published proposals for mobility support on IP layer are presented. The most important mechanism is the IP Mobility support for IPv4 specified in RFC 3344. This protocol was initially published by Charles E. Perkins and still forms the basis for all other protocols.

Mobile networks of the 4th generation will exhibit a need for mobility support even for QoS demanding devices and applications. It is clear, that the basic Mobile IP protocol does not support such high demands. It was designed to allow mobility at all. Therefore, extensions to the basic approach are definitely needed.

Nevertheless, there is no single proposal that can provide all the necessary capabilities. Some extensions deal with the optimization of Micro-Mobility, while others improve the overall delays, e.g. by avoiding triangle routing. Yet others discuss the important issue of authentication and authorization when mobility occurs. None of these proposals considering AAA has been discussed in this section, which does not mean that they are of less importance.

However, there is a large number of publications that are still unmentioned. The IPv6 protocol, for example, provides support for extension headers, which again allow to integrate mobility support directly in the IP protocol itself. All of the proposals discussed above can be found in the IPv6 domain as well. In addition, IPv6 allows for other extensions [CH02]. The support for QoS within the IP protocol is just one issue. Therefore, other methods to improve the performance of handovers become possible.

All of these factors clearly show that a myriad of different proposals and extensions to the basic Mobile IP protocol exist. On the other hand, the discussion about 4G networks just started and the architecture is not completely clear yet.

This all adds to the uncertainty of which protocol is the best candidate for future mobile networks. Clearly no recommendation can be given so far. Extensive studies of the various protocols are still necessary to get a better understanding of the differences and their suitability in different potential architectures.

7 Conclusion and Outlook

> The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' (I found it!) but 'That's funny ...'. Isaac Asimov (1920 - 1992)

Wireless networks of the 4th generation are both, a great opportunity and a big challenge. The integration of heterogeneous networks into one single system allows to exploit the individual strengths of each technology but also increases the complexity. The industry has long identified the potentials and started research projects to study the feasibility. This fact is reflected by the topics seen at international research conferences that started to pick up issues of 4G networks into their schedules lately.

However, research in the area of 4G networks has just started. Discussions indicate that there might be potential for a competition between Wireless LAN based networks and UMTS systems in the future. Nevertheless, these discussions are mostly based on visions and not on a solid basis of research results. Our studies are intended to shed light on one specific topic. The question was whether the Wireless LAN protocol is capable of simultaneously transporting data traffic of different types while still keeping the user-experienced Quality of Service at an acceptable or necessary level. This is an important factor, since it is essential if Wireless LAN networks can be integrated into today's legacy systems where these kinds of data traffic occur and are supported.

We started our discussion with an overview of the history of wireless systems in Chapter 2. It shows the evolutionary development of mobile communication up to date. This is followed by a summary of the Wireless LAN standard and the various standard extensions. In addition to the basic standard IEEE 802.11, which all Wireless LAN devices are based on, a total of 19 different standard extensions are listed. Some of them reached the state of an official supplement, while others are still in the process of being developed. This clearly indicates that the development of the Wireless LAN protocol is by far not finished, yet. As the discussion about the future potential of Wireless LAN continues while research results frequently show the advantages or drawbacks of various WLAN features, the protocol is further enhanced to overcome any appearing problems; Wireless LAN evolves evolutionary.

The most important standard supplement for our studies is the IEEE 802.11e *Medium Access Control (MAC) Quality of Service (QoS) Enhancements*. It defines extensions to the MAC protocol itself. The goal is to differentiate traffic in a way that a prioritization of QoS demanding traffic can be achieved over best-effort traffic. However, the exact properties of the different traffic types have to be taken into account in order to receive valid and significant results. Chapter 3 takes a closer look at these traffic types. It discusses the File Transfer Protocol (FTP) and the Hypertext Transfer Protocol (HTTP), and explains the way voice and video data is transported. This includes a discussion of voice and video codecs as well as of the User Datagram Protocol (UDP), Transport Control Protocol (TCP), and Real-Time Protocol (RTP).

Chapter 3 then explains the differences in quality assessments that are necessary depending on the type of traffic that is considered. While a Web user mostly cares about the amount of time that he has to wait until the download of the requested page is finished, it is much more complicated to judge whether a voice or video user is satisfied with the received Quality of Service. Different approaches as found in the literature are explained. The chapter is finished by a summary of the traffic types that are considered in the studies presented later.

The Physical layer specifics of the Wireless LAN protocol are explained in Chapter 4. The different modulation techniques such as BPSK, QPSK, CCK, and OFDM are presented in detail and their properties and implications on the performance of the wireless channel are discussed. The chapter is continued with a summary of the important factors that have to be taken into account when analyzing a wireless channel. This includes issues such as fading, path loss, and the ways in which bit error rates or packet error rates can be calculated. The chapter is concluded by an overview of the parameters used to retrieve the results of later chapters.

The analysis of the Wireless LAN Medium Access Control protocol and its capability to support QoS demanding traffic and applications is presented in Chapter 5. In the first part, the different MAC protocols as defined in the IEEE 802.11 standard and its extensions are explained in detail. This includes the basic DCF operation mode, with the CSMA/CA protocol, as well as further enhancements, such as the polling mechanism specified in the PCF mode and finally the HCF mode, with its different access mechanisms EDCA and HCCA.

The results for the MAC protocol performance evaluation are presented in Section 5.3. It starts with simple single cell scenarios and basic performance studies of the CSMA/CA protocol in the case where only Web users are present. After that, an extensive study of the Hidden Node problem and the RTS/CTS mechanism follows. It is shown that the RTS/CTS mechanism can not improve the system performance due the fact that it increases the overhead of the protocol.

In the following, overlapping and co-located cells are studied. It is expected that these cases will occur frequently in large-scale environments as they are expected when 4G systems are deployed. The Physical layer restrictions that only allow three non-overlapping channels, lead to scenarios where the coverage areas of Wireless LAN access points overlap both in coverage and in frequency, such that great amounts of interference are the consequence. Our studies aim at the question of how well the different MAC protocols can deal with such situations while the performance is kept at a high level and QoS demanding applications and traffic has to be supported.

Section 5.3.2 shows that the pure DCF mode of operation is not capable to cope with the problems arising in overlapping or co-located cells, not even in the case where only best-effort traffic is studied. However, it is shown that if a prioritization scheme based on the Contention Window sizes, and thus on the average number of back-off slots is introduced, the unfairness of the CSMA/CA protocol can be eased in the studied scenarios. However, it is pointed out that the DCF mode, as defined in the standard, does not allow to perform such a prioritization scheme. Therefore, it can be concluded that it is simply not capable of providing fairness and QoS support in overlapping and co-located cells.

In the next step, the HCF mode with its QoS enabling functionality for single cell scenarios is studied. This IEEE 802.11e enhancement of the basic CSMA/CA protocol allows to perform a prioritization based on the Contention Window size. Therefore, it seems appropriate to perform the needed prioritization for overlapping and co-located cells. A large number of different simulation scenarios are studied in Section 5.3.3. Different mixtures of Wireless LAN clients performing voice, video, and best-effort traffic in the overlapping and co-located cell scenarios are evaluated. It turns out that the prioritization based on the Contention Window size is adequate to support any kind of traffic mixture. Depending on the scenario, different sets of prioritization have to be chosen in order to provide the necessary QoS level for real-time demanding applications while still keeping the performance experienced by best-effort users at a high level. Finally, it is shown that it is possible to simultaneously perform voice, video, and best-effort applications in any of the scenarios and still keep the quality for any of the three at an acceptable level.

It is clear that the system can only provide QoS service to its associated clients, as long as the system does not experience an overload situation. In order to prevent such situations, however, an access control mechanism has to be applied. Nevertheless, this was not the topic of our studies, such that no access control mechanisms have been implemented. It was merely shown that if the system load is kept at an appropriately low level, the HCF MAC mechanism can provide adequate prioritization to support mixtures of real-time and best effort traffic even in the challenging scenarios of overlapping and co-located cells.

Another important issue when studying wireless systems and their capability of supporting QoS traffic is the topic of handover. Therefore, Wireless LAN stations roaming between different access points are studied extensively in Section 5.4. The handover procedure consists of three different tasks. First, scanning has to be performed. It identifies potential candidate access points that the Wireless LAN station can associate with. Then, association or reassociation has to be performed. After this stage, the station is connected to the new access point. Finally, the authentication mechanism has to make sure that the client is allowed to use the new access point.

It is identified that scanning is the most critical part of a handover in terms of performance. A number of different scanning mechanisms are defined in the standards, such as active and passive scanning as well as variations of these two. Our studies show that not all of these candidates perform equally well in the complex scenarios considered in our simulations. However, it can be concluded that the fast active scanning mechanism performs well enough to allow handovers of stations performing real-time applications even in overlapping and co-located cells.

Summarizing Chapter 5 we can conclude that the IEEE 802.11e standard with its HCF mode of operation is well suited to support a mixture of real-time and best-effort traffic in complex Wireless LAN scenarios. It is important to choose the right parameters, and an appropriate admission control mechanism to avoid system overload is required, but Wireless LAN in a 4G environment is possible if it comes to the MAC protocol.

It is important to realize, that there exist some more requirements for the handover in large-scale Wireless LAN environments. Section 5.4 studied the requirements of a handover within a single IP subnet. Here, the station roams to a new access point, but it can keep up its communication to other nodes because a change of the IP address was not necessary. However, in large networks, handovers occur between different IP subnets. In such a case, the IP routing within the Internet requires a change of the IP address of the client as well. Traditionally this is not possible seamlessly without additional changes to the protocols. Otherwise, all the applications have to be informed of a change of IP address, which usually requires a complete restart.

Therefore, the Mobile IP protocol was proposed. It allows for seamless handovers. However, the initial draft of the protocol does not deal with the problem of delays during the handover process. In Chapter 6 an overview of the basic Mobile IP protocol is given and some performance enhancing extensions to the protocol are discussed. Due to large amount of proposed mechanisms, it is impossible to give a complete overview of the topic in this work. The point merely was to give the reader a feeling of what the different proposals aim at. At the end of the chapter we discuss the fact that it is still uncertain which protocol is the best candidate for future mobile networks. No recommendations can be given here, but more research work is needed.

Considering the results of our studies, we can conclude that Wireless LAN has a great potential in the context of 4G networks. There are still many open issues that have to be clarified before a final answer can be given, but it is certain that the future Wireless LAN MAC protocols can provide the necessary QoS levels. From a technical perspective it is possible that mobile networks of the 4th generation including the Wireless LAN technology can become a reality.

Bibliography

[AHP$^+$01] M. Annoni, R. Hancock, T. Paila, E. Scarrone, R. Toenjes, L. Dell'Uomo, and D. Wisely. Radio Access Networks beyond the 3rd Generation: A first Comparison of Architectures of 4 IST Projects. *Mobile Summit 2001*, 2001.

[AMC$^+$99] I.F. Akyildiz, J. McNair, L. Carrasco, R. Puigjaner, and Y. Yesha. Medium Access Control Protocols for Multimedia Traffic in Wireless Networks. *IEEE Network Magazine*, 1999.

[BCG02] R. Bruno, M. Conti, and E. Gregori. IEEE 802.11 Optimal Performances: RTS/CTS Mechanism vs. Basic Access. *In Proceedings of the 13th IEEE Intl. Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2002.

[BS03] S. Black and H. Sinivaara. IEEE P802.11 Wireless LANs - Revised Proposal for the Distribution of Neighbourhood BSS Information. Technical report, Nokia, 2003. IEEE 802.11-03/580r1.

[BVC01] M. Barry, A. Veres, and A. Campbell. Distributed Control

Algorithms for Service Differentiation in Wireless Packet Networks. In *Proceedings of Infocom 2001*, 2001.

[CGKT02] D. Chen, S. Garg, M. Kappes, and K. Trivedi. Supporting VBR VoIP Traffic in IEEE 802.11 WLAN in PCF Mode. Technical report, Avaya Labs Research, 233 Mount Airy Road, Basking Ridge, New Jersey 07920, 2002.

[CGL00] A. Chandra, V. Gummalla, and J. O. Limb. Wireless Medium Access Control Protocols. *IEEE Communications Surveys & Tutorials*, Second Quarter 2000.

[CH02] X. P. Costa and H. Hartenstein. A Simulation Study on the Performance of Mobile IPv6 in a WLAN-based Cellular Network. *Source Computer Networks: The International Journal of Computer and Telecommunications Networking*, 2002.

[Cis04] Cisco Systems, Inc. Understanding codecs: Complexity, hardware support, mos, and negotiation. Technical report, Document ID: 14069, Cisco Systems, Inc., 2004.

[Cov01] P. Coverdale. Multimedia QoS requirements from a user perspective, 2001. Nortel Networks and ITU-T Study Group 12.

[CSL02] A. Chindapol, A. Stephens, and J. Lansford. IEEE P802.15 Wireless Personal Area Networks - Change request for 802.15.2 Recommended Practice PHY test. Technical report, IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), 2002. IEEE P802.15-02/069r0.

[CWKS97] B. Crow, I. Widjaja, J. G. Kim, and P. Sakai. Investigation of the IEEE 802.11 Medium Access Control (MAC)

Sublayer Functions. In *Proceedings of the INFOCOM '97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, 1997.

[Eur94] European digital cellular telecommunications system (Phase 2); Source: SMG 2. Full rate speech processing functions (GSM 06.01). Technical report, European Telecommunication Standards Institute, 1994.

[FSR04] F. Fitzek, P. Seeling, and M. Reisslein. *Video Streaming in Wireless Internet.* Wireless Internet: Technologies and Applications Series: Electrical Engineering & Applied Signal Processing Series, CRC Press, 2004. To be published.

[Gas02] M. Gast. *802.11 Wireless Networks: The Definitive Guide.* O'Reilly Networking; 1 edition, 2002.

[Gei01] J. Geier. *Wireless LANs - Implementing High Performance IEEE 802.11 Networks.* SAMS, 2 edition, 2001.

[Gib97] J. D. Gibson. *The Communications Handbook.* Electrical engineering handbook series, CRC Press, Inc., 1997.

[GJP04] Eva Gustafsson, Annika Jonsson, and Charles E. Perkins. Mobile IPv4 Regional Registration, 2004. draft-ietf-mobileip-reg-tunnel-09.txt, Mobile IP Working Group, INTERNET DRAFT.

[GK03] S. Garg and M. Kappes. An experimental study of throughput for UDP and VoIP traffic in IEEE 802.11b networks. In *Proceedings of the Wireless Communications and Networking, 2003. WCNC 2003*, 2003.

[GW99] A. Ganz and K. Wongthavarawat. IEEE 802.11 Wireless LAN Association Procedure for Multimedia Applications.

In Proceedings of the IEEE Military Communications Conference, October 1999.

[GZ03] D. Gu and J. Zhang. QoS Enhancements in IEEE 802.11 Wireless Local Area Networks. *IEEE Communications Magazine*, June 2003.

[Hat80] M. Hata. Empirical formulae for propagation loss in land mobile radio services. In *Proc. of IEEE VTC*, pages 317–325, 1980.

[Hec03a] K. Heck. Web Traffic Performance in Wireless LAN Hot Spots. *In Proceedings of the Fourth Internatinal Conference on 3G Mobile Communication Technologies (3G 2003)*, June 2003.

[Hec03b] K. Heck. Wireless LAN Performance in Overlapping Cells. *In Proceedings of the 58th Vehicular Technology Conference (VTC Fall 2003)*, October 2003.

[HPW04] K. Heck, R. Pries, and T. Wirth. Simulative Study of the RTS/CTS Solution to the Hidden Node Problem in Infrastructure Wireless LANs. *In Proceedings of the World Wireless Congress (WWC 2004)*, May 2004.

[HRW03] C. Hoene, B. Rathke, and A. Wolisz. On the Importance of a VoIP Packet. *In Proceedings of ISCA Tutorial and Research Workshop on the Auditory Quality of Systems*, 2003.

[HSL02] Klaus Heck, Dirk Staehle, and Kenji Leibnitz. Diversity Effects on the Soft Handover Gain in UMTS networks. In *Proceedings of the IEEE Vehicular Technology Conference*, Vancouver, Canada, September 2002.

[IEE98] IEEE. IEEE Standard for Information Technology - Telecommunications and information exchange between systems - Local and Metropolitan networks - Specific requirements - Part 3: Media Access Control (MAC) Bridges, 1998. IEEE 802.1d-1998.

[IEE99a] IEEE. 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, 1999. ANSI/IEEE 802.11, 1999 Edition.

[IEE99b] IEEE. IEEE Standard for Information Technology - Telecommunications and information exchange between systems - Local and Metropolitan networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher speed Physical Layer (PHY) extension in the 2.4 Ghz band, 1999. IEEE Std 802.11b-1999.

[IEE99c] IEEE. 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, 1999. IEEE Std 802.11a-1999.

[IEE03a] IEEE. 802.11F-2003 IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation, 2003. IEEE Std 802.11f-2003.

[IEE03b] IEEE. 802.16A-2003 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, 2003. IEEE Std 802.16A-2003.

[IEE03c] IEEE. IEEE Standard for Information Technology - Telecommunications and information exchange between systems - Local and Metropolitan 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, June 2003. IEEE Std 802.11g-2003.

[IEE03d] IEEE. IEEE Standard for Information Technology - Telecommunications and information exchange between systems - Local and Metropolitan networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), June 2003. IEEE 802.11e/D4.4-2003.

[IEE03e] IEEE. IEEE Standard for Information Technology - Telecommunications and information exchange between systems - Local and Metropolitan networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), November 2003. IEEE 802.11e/D5.2-2003.

[Inf01a] Information Society Technologies (IST). BRAIN Architecture Specifications and Models, BRAIN Functionality and Protocol Specification, 2001. IST-1999-10050 BRAIN D2.2.

[Inf01b] Information Society Technologies (IST). Moby Dick - Mobility and Differentiated Services in a Future IP Network - Dissemination and Use Plan, 2001. IST-2000-25394 D0601.

[Inf02] Information Society Technologies (IST). Wireless IP Network as a Generic Platform for Location Aware Service Support, 2002. IST-1999-10699 WINE GLASS - D016 - Final Project Report.

[Inf03] Information Society Technologies (IST). EVOLUTE - Final Evaluation Report, 2003. IST-2001-32449 D4.2.

[IT93a] ITU-T. Pulse Code Modulation (PCM), 1993. ITU-T Recommendation G.711.

[IT93b] ITU-T. Video Codec for Audiovisual Services at p x 64kbit/s, 1993. ITU-T Recommendation H.261.

[IT95] ITU-T. Video Coding for Low Bitrate Communication (TMN5), 1995. ITU-T Recommendation H.263.

[IT96a] ITU-T. 40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM), 1996. ITU-T Recommendation G.723.1.

[IT96b] ITU-T. Coding of Speech at 8 kbit/s using Conjugate Structure Algebraic Code-Excited Linear Prediction (CS-ACELP), 1996. ITU-T Recommendation G.729.

[IT96c] ITU-T. Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, 1996. ITU-T Recommendation G.723.1.

[IT96d] ITU-T. Series P: Telephone Transmission Quality, Methods for Subjective Determination of Transmission Quality, 1996. ITU-T Recommendation P.800.

[IT96e] ITU-T. Subjective Performance Assessment of Telephone-band and Wideband Digital Codecs, 1996. ITU-T Recommendation P.830.

[IT98] ITU-T. Vocabulary of terms on telephone transmission quality and telephone sets, 1998. ITU-T Recommendation P.10.

[IT01] ITU-T. Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs, 2001. ITU-T Recommendation P.862.

[IT02] ITU-T. Advanced Video Coding, 2002. ITU-T Recommendation H.264, ISO/IEC 11496-10, Final Committee Draft, Document JVT-Eo22.

[JWKZ03] M. Jeong, F. Watanabe, T. Kawahara, and Z. Zhong. Proposed Text for Fast Active Scan. Technical report, Contribution to IEEE802.11-03/623r0, 2003.

[Kab03] P. Kabal. ITU-T G.723.1 Speech Coder: A Matlab Implementation. Technical report, Department of Electrical & Computer Engineering, TSP Labs, McGill University, 2003.

[KEW00] A. Köpsel, J. Ebert, and A. Wolisz. A Performance Comparison of Point and Distributed Coordination Function of an IEEE 802.11 WLAN in the Presence of Real-Time Requirements. *In Proceedings of the 7th. Intl. Workshop on Mobile Multimedia Communications (MoMuC)*, 2000.

[KW04] A. Köpsel and A. Wolisz. Voice Transmission in an IEEE 802.11 WLAN Based Access Network. *In Proceedings of the Cooperative Internet Computing (CIC)*, 2004.

[LAS01a] A. Lindgren, A. Almquist, and O. Scheln. Evaluation of Quality of Service Schemes for IEEE 802.11 Wireless LANs. In *Proceedings of the 26th Annual IEEE Conference on Local Computer Networks (LCN 2001)*, 2001.

[LAS01b] A. Lindgren, A. Almquist, and O. Scheln. Quality of Service Schemes for IEEE 802.11 - A Simulation Study. In *Proceedings of the Ninth International Workshop on Quality of Service (IWQoS 2001)*, 2001.

[LGT01] R. Langar, G. Le Grand, and S. Tohme. Micro Mobile MPLS Protocol in Next Generation Wireless Access Networks. *In Proceedings of the WoWMoM*, 2001.

[LSDS01] J. Lansford, A. Stephens, R. E. Van Dyck, and A. Soltanian. IEEE P802.15 Wireless Personal Area Networks - (Combined) Mobilian and NIST Text for Clause 6. Technical report, IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), 2001. IEEE P802.15-01418r0.

[Luc98] Lucent Technologies Inc. Roaming With WaveLAN/IEEE 802.11. Technical report, Lucent Technologies Inc, WaveLAN Technical Bulletin 021/A, 1998.

[Mal04] K. El Malki. Low Latency Handoffs in Mobile IPv4, 2004. draft-ietf-mobileip-lowlatency-handoffs-v4-09.txt, Network Working Group, INTERNET DRAFT.

[MCM$^+$02] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor. IEEE 802.11e Wireless Lan for Quality of Service. In *Proceedings of the European Wireless*, Florence, Italy, 2002.

[MSA02] A. Mishra, M. Shin, and W. Arbaugh. An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process. Technical Report CS-TR-4395, University of Maryland Department of Computer Science, September 2002.

[NH95] A. N. Netravali and B. G. Haskell. *Digital Pictures - Representation, Compression and Standards.* New York: Plenum, 1995.

[Oku68] Y. Okumura. Field Strength and its Variability in VHF and UHF Land Mobile Radio Service. In *Rev. Elec. Comm. Lab.*, pages 825–873, 1968.

[OP99] B. O'Hara and A. Petrick. *IEEE 802.11 Handbook: A Designer's Companion.* Standards Information Network, IEEE Press, 1999.

[PC02] S. Pack and Y. Choi. Pre-Authenticated Fast Handoff in a Public Wireless LAN based on IEEE 802.1x Model, 2002. IFIP TC6 Personal Wireless Communications.

[Per96] Charles E. Perkins. IP Mobility Support, 1996. Network Working Group, Request for Comments, 2002.

[Per02] Charles E. Perkins. IP Mobility Support for IPv4, 2002. Network Working Group, Request for Comments, 3344.

[PF03] C. Peikari and S. Fogie. *Maximum Wireless Security - An Insider's Guide to Protecting Your Network*. Sams Publishing, 2003.

[PH04] R. Pries and K. Heck. Performance Comparison of Handover Mechanisms in Wireless LAN Networks. In *Proceedings of the Australian Telecommunications, Networks and Applications Conference (ATNAC)*, Sydney, Australia, 2004.

[PHWTG04] R. Pries, K. Heck, T. Wirth, and P. Tran-Gia. Robustness Analysis of the Wireless LAN MAC Protocols with QoS Support. Technical Report 346, University of Würzburg, 11 2004.

[PJ02] Charles E. Perkins and David B. Johnson. Route Optimization in Mobile IP, 2002. draft-ietf-mobileip-optim-12.txt, Mobile IP Working Group, INTERNET DRAFT.

[Pos80] J. Postel. User Datagram Protocol (UDP), 1980. Information Sciences Institute (ISI), Request for Comments, 768.

[Pos81] J. Postel. Transmission Control Protocol (TCP), 1981. Information Sciences Institute (ISI), Request for Comments, 793.

[PP01] N. Prasad and A. Prasad. *WLAN Systems and Wireless IP for Next Generation Communications.* Artech House universal personal communcations series, 2001.

[PR85] J. Postel and J. Reynolds. FILE TRANSFER PROTOCOL (FTP), 1985. Information Sciences Institute (ISI), Request for Comments, 959.

[RAHE01] R. S. Ranasinghe, L. L. H. Andrew, D. A. Hayes, and D. Everitt. Scheduling Disciplines for Multimedia WLANs: Embedded Round Robin and Wireless Dual Queue. In *Proceedings of the IEEE International Conference on Communications (ICC)*, 2001.

[Rap96] T. S. Rappaport. *Wireless Communications - Principles & Practice*. Prentice Hall, Inc., 1996.

[rGPP02a] 3rd Generation Partnership Project. Technical Specification Group Services and System Aspects; 3GPP system to Wireless Local Area Network (WLAN) interworking; Functional and architectureal definition (Release 6), 2002. 3GPP TR 23.934 V1.0.0.

[rGPP02b] 3rd Generation Partnership Project. Technical Specification Group Services and System Aspects; Feasibility study on 3GPP system to Wireless Local Area Network (WLAN) interworking (Release 6), 2002. 3GPP TR 22.934 V6.1.0.

[rGPP02c] 3rd Generation Partnership Project. Technical Specification Group Services and System Aspects; Wireless Local Area Network (WLAN) interworking; Security (Release 6), 2002. 3GPP TS 33.cde V0.1.0.

[rGPP03] 3rd Generation Partnership Project. Technical Specification Group Services and System Aspects; 3GPP system to Wireless Local Area Network (WLAN) interworking; System Description (Release 6), 2003. 3GPP TR 23.234 V1.3.0.

[Ric99] I. Richardson. *Video Coding for Reliable Communications*. PhD thesis, The Robert Gordon University, 1999.

[RJ91] M. Rabbani and P. W. Jones. *Digital Image Compression Techniques*. Tutorial Text Vol. TT07, Society of Photo-Optical Instrumentation Engineers (SPIE), 1991.

[RTFK01] Z. Ren, C. Tham, C. Foo, and C. Ko. Integration of Mobile IP and Multi-Protocol Label Switching. *ICC 2001 - IEEE International Conference on Communications*, June 2001.

[RVS$^+$99] R. Ramjee, K. Varadhan, L. Salgarelli, S. Thuel, S. Y. Wang, and T. La Porta. HAWAII: A Domain-based Approach for Supporting Mobility in Wide-area Wireless Networks . *In Proceedings of the International Conference on Network Protocols (ICNP)*, 1999.

[SCFJ03] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications, 2003. Network Working Group, Request for Comments, 3550.

[Sch03] J. Schiller. *Mobilkommunikation*. Pearson Studium, Addison-Wesley, 2003.

[SK96] J. Sobrinho and A. Krishnakumar. Real-Time Traffic over the IEEE 802.11 Medium Access Control Layer. *Bell Labs Technical Journal*, 1996.

[SK98] M. Stemm and R. Katz. Vertical Handoffs in Wireless Overlay Networks. *Mobile Networks and Applications*, 1998.

[Ste94] W. Stevens. *TCP/IP Illustrated - Volume 1: The Protocols*. Addison-Wesley, 1994.

[TGSL01] P. Tran-Gia, D. Staehle, and K. Leibnitz. Source Traffic Modeling of Wireless Applications. *International Journal of Electronics and Communications (AE)*, 55, 2001.

[Tyc02] O. Tyce. Handover Optimisation. Technical report, School of Cognitive and Computing Sciences, University of Sussex, 2002.

[Val99] A. G. Valko. Cellular IP: A New Approach to Internet Host Mobility. *ACM SIGCOMM Computer Communication Review*, 1999.

[VBG00] N. H. Vaidya, P. Bahl, and S. Gupta. Distributed Fair Scheduling in a Wireless LAN. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, 2000.

[VC03] A. Velayutham and J. Morris Chang. An Enhanced Alternative to the IEEE 802.11e MAC Scheme. Technical report, Department of Computer Science, Iowa State University, 2003.

[VCBS01] A. Veres, A. T. Campbell, M. Barry, and L.-H. Sun. Supporting Service Differentiation in Wireless Packet Networks using Distributed Control. *IEEE Journal of Selected Areas in Communications (J-SAC)*, 2001.

[XS02] S. Xu and T. Saadawi. Revealing the Problems with 802.11 Medium Access Control Protocol in Multi-hop Wireless Ad Hoc Networks. *Computer Networks, Vol. 38, pp. 531-548*, 2002.

[YM01] T. Yang and D. Makrakis. Hierarchical Mobile MPLS: Supporting Delay Sensitive Applications Over Wireless Internet. *In Proceedings of the International Conferences on Info-tech and Info-net (ICII 2001)*, October 2001.

[ZJD03] W. Zhang, J. Jaehnert, and K. Dolzer. Design and Evaluation of a Handover Decision Strategy for 4th Generation Mobile Networks. *In Proceedings of the 57th Vehicular Technology Conference (VTC Spring 2003)*, 2003.

Glossary

16-QAM 16-Quadrature Amplitude Modulation

1G First generation mobile networks

2.5G Second and a half generation mobile networks

2G Second generation mobile networks

3G Third generation mobile networks

3GPP The Third Generation Partnership Project

3GPP2 The Third Generation Partnership Project 2

4G Forth generation mobile networks

64-QAM 64-Quadrature Amplitude Modulation

AAA Authentication, Authorization, and Accounting

AC Access Category

ACELP Algebraic Code Excited Linear Prediction

ADPCM Adaptive Differential Pulse Code Modulation

AIFS Arbitration Interframe Space

AMPS Advanced Mobile Phone Service

APME Access Point Management Entity

ARP Address Resolution Protocol

B-Frame Bidirectional predicted frame

BCC Binary Convolutional Code

BER Bit error rate

BPSK Binary Phase Shift Keying

BRAN Broadband Radio Access Networks

BSS Basic Service Set

CAF Channel Access Function

CAP Controlled Access Phase

CCK Complementary Code Keying

CDMA Code Division Multiple Access

cdmaOne CDMA technology based on IS-95

CFP Contention-Free Period

CIF Common Interframe Format

CN Correspondent Node

CS-ACELP Conjugate Structure Algebraic Code Exited Linear Prediction

CSMA/CA Carrier-Sense Multiple Access with Collision Avoidance

CTS Clear To Send control frame

D-AMPS Digital Advanced Mobile Phone Service

dB decibel

dBi Decibel referenced to an isotropic radiator

dBm Decibels referenced to 1 milliwatt

DBPSK Differential Binary Phase Shift Keying

DCF Distributed Coordination Function

DCS1800 Digital Cellular System - 1800

DCT Discrete Cosine Transform

DHCP Dynamic Host Configuration Protocol

DIFS Distributed (Coordination Function) Interframe Space

DQPSK Differential Quadrature Phase Shift Keying

DS Distribution System

DSSS Direct-Sequence Spread Spectrum

EDCA Enhanced Distributed Channel Access

EIA Electronic Industries Alliance

EIFS Extended Interframe Space

ERP Effective Radiated Power

ESS Extended Service Set

ESS Extended Service Set

ETSI European Telecommunications Standards Institute

FDA Foreign Domain Agent

FDD Frequency Division Duplex

FDMA Frequency Division Multiple Access

FFT Fast Fourier Transform

FHSS Frequency Hopping Spread Spectrum

FOMA Freedom Of Mobile multimedia Access

FTP File Transfer Protocol

GFSK Gaussian Frequency Shift Keying

GMSK Gaussian Minimum Shift Keying

GoP Group of Pictures

GPRS General Packet Radio Service

GSM Global System for Mobile Communications

HC Hybrid Coordinator

HCCA HCF Controlled Channel Access

HCF Hybrid Coordination Function

HIPERLAN HIgh PErformance Radio LAN

HTTP Hypertext Transfer Protocol

I-Frame Intra-coded frame

IAPP Inter Access Point Protocol

IBSS Independent Basic Service Set

IEEE Institute of Electrical and Electronics Engineers

IETF Internet Engineering Task Force

IFFT Inverse Fast Fourier Transform

IP Internet Protocol

IR Infrared Physical layer

IS Interim Standard

ISO International Organization for Standardization

ITU International Telecommunication Union

ITU-T ITU Telecommunication Standardization Section

JTACS Japanese Total Access Communication System

KBps Kilo byte per second

Kbps Kilo bit per second

LAN Local Area Network

LLC Logical Link Control

LOS Line Of Sight

MAC Medium Access Control

Mcps Mega chips per second

MIMO Multiple-Input-Multiple-Output

MLME MAC Layer Management Entity

MOS Mean Opinion Score

MP-MLQ Multi-Pulse Maximum Likelihood Quantization

MPLS Multiprotocol Label Switching

NAMPS Narrowband Advanced Mobile Phone Service

NAT Network Address Translation

NAV Network Allocation Vector

NLOS Non Line Of Sight

NMT Nordic Mobile Telephone

NTACS Narrowband Total Access Communication System

NTT Nippon Telephone and Telegraph

OFDM Orthogonal Frequency Division Multiplex

OSI Open Systems Interconnect

P-Frame Predicted frame

PAN Personal Area Network

PBCC Packet Binary Convolutional Coding

PC Personal Computer

PC Point coordinator

PCF Point Coordination Function

PCM Pulse Code Modulation

PCS Personal Communication Services

PDA Personal Digital Assistant

PESQ Perceptual Evaluation of Speech Quality

PHY Physical layer

PIFS Point (Coordination Function) Interframe Space

PLME Physical Layer Management Entity

PSNR Peak Signal to Noise Ratio

QCIF Quarter Common Interframe Format

QoS Quality of Service

QPSK Quadrature Phase Shift Keying

QSTA QoS stations

RADIUS Remote Authentication Dial-In User Service

RTP Real-time Transport Protocol

RTS Request To Send control frame

SAP Service Access Point

SER Symbol error rate

SIFS Short Interframe Space

SNR Signal to Noise Ratio

SSID Service Set ID

TACS Total Access Communication System

TBTT Target Beacon Transmission Time

TCP Transmission Control Protocol

TDMA Time Division Multiple Access

TIA Telecommunications Industry Association

TXOP Tansmission Opportunity

UDP User Datagram Protocol

UMTS Universal Mobile Telecommunication System

UP User Priority

WCDMA Wideband Code Division Multiple Access

WDS Wireless Distribution System

WEP Wired Equivalent Privacy

WiFi Wireless Fidelity

WiMAX Worldwide Interoperability for Microwave Access

WISP Wireless Internet Service Providers

WLAN Wireless Local Area Network

WMPLS Wireless Multiprotocol Label Switching

WN Who is New control frame

WWW World Wide Web

Index

ISSN 1432-8801