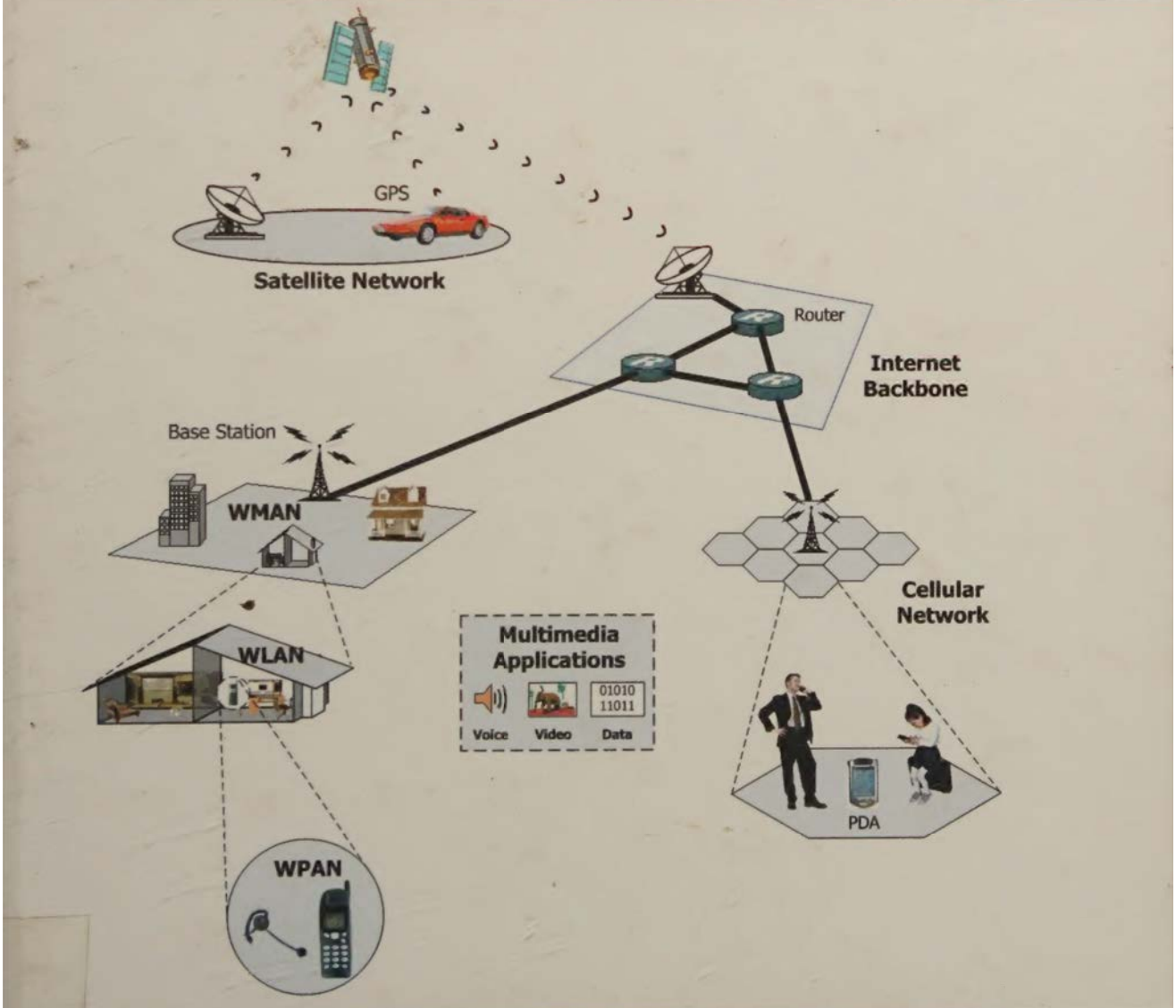
MULTIMEDIA
WIRELESS NETWORKS
Technologies, Standards, and QoS

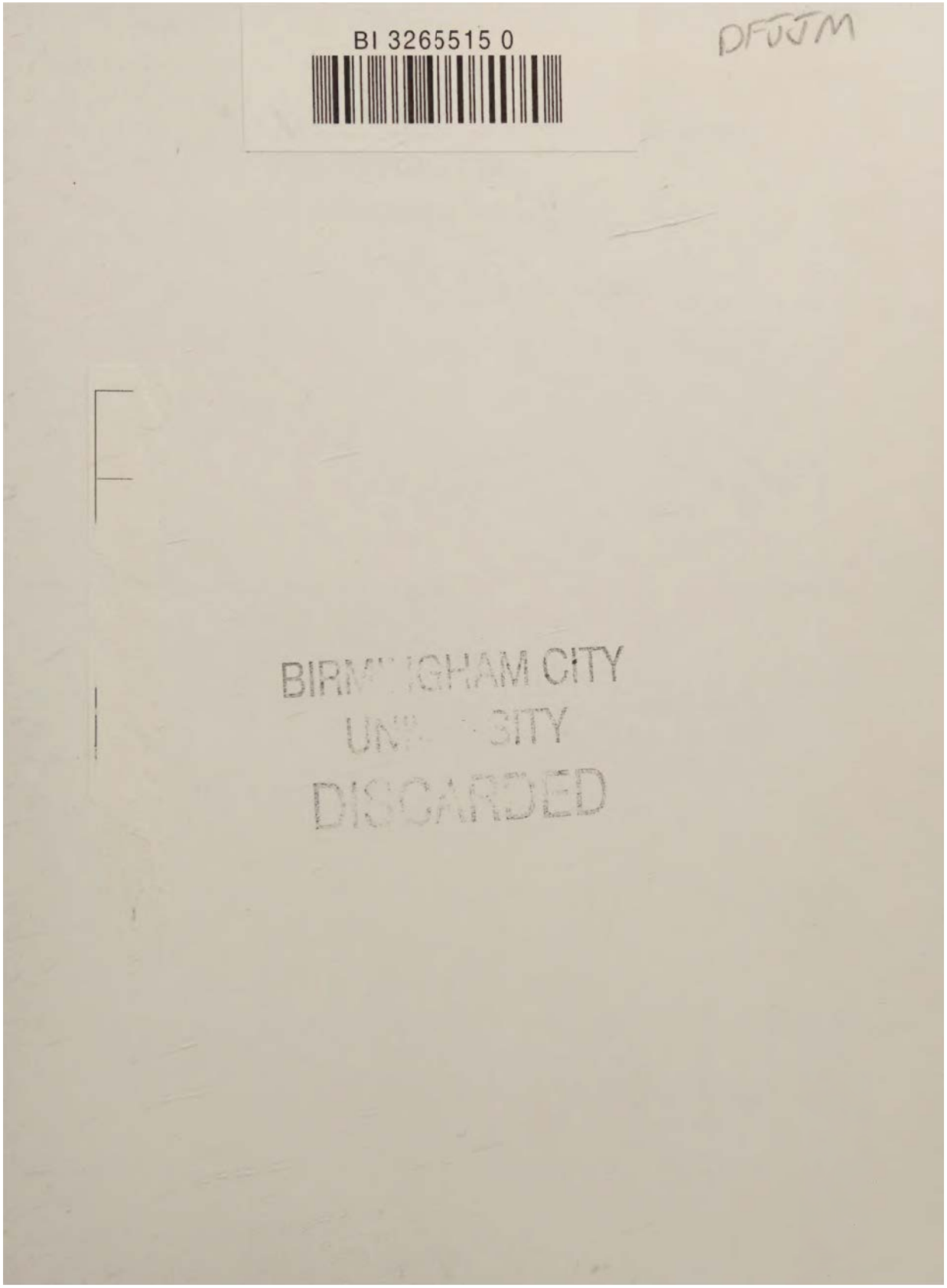


Aura Ganz, Zvi Ganz, and Kitti Wongthavarawat

Prentice Hall Communications Engineering and Emerging Technologies Series
Theodore S. Rappaport, Series Editor

SAMSUNG 1019




BI 3265515 0
DFUJM
BIRMINGHAM CITY
UNIVERSITY
DISCARDED



MULTIMEDIA WIRELESS NETWORKS

Technologies, Standards, and QoS

Prentice Hall PTR Communications Engineering and Emerging Technologies Series

Theodore S. Rappaport, *Series Editor*

DOSTERT *Powerline Communications*

DURGIN *Space–Time Wireless Channels*

GANZ, GANZ, AND WONGTHAVARAWAT *Multimedia Wireless Networks: Technologies, Standards, and QoS*

GARG *Wireless Network Evolution: 2G to 3G*

GARG *IS-95 CDMA and cdma2000: Cellular/PCS Systems Implementation*

GARG & WILKES *Principles and Applications of GSM*

HAĆ *Multimedia Applications Support for Wireless ATM Networks*

KIM *Handbook of CDMA System Design, Engineering, and Optimization*

LIBERTI & RAPPAPORT *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*

PAHLAVAN & KRISHNAMURTHY *Principles of Wireless Networks: A Unified Approach*

RAPPAPORT *Wireless Communications: Principles and Practice, Second Edition*

RAZAVI *RF Microelectronics*

REED *Software Radio: A Modern Approach to Radio Engineering*

STARR, CIOFFI & SILVERMAN *Understanding Digital Subscriber Line Technology*

STARR, SORBARA, CIOFFI & SILVERMAN *DSL Advances*

TRANTER, SHANMUGAN, RAPPAPORT & KOSBAR *Principles of Communication Systems Simulation with Wireless Applications*

MULTIMEDIA WIRELESS NETWORKS

Technologies, Standards, and QoS

Aura Ganz
Zvi Ganz
Kitti Wongthavarawat



PRENTICE HALL PTR
UPPER SADDLE RIVER, NJ 07458
WWW.PHPTR.COM



Library of Congress Cataloging-in-Publication Data

A catalog record for this book can be obtained from the Library of Congress

Publisher: *Bernard Goodwin*
Editorial/production supervision: *Nicholas Radhuber*
Cover design director: *Jerry Votta*
Cover design: *Nina Scuderi*
Manufacturing manager: *Maura Zaldivar*
Editorial assistant: *Michelle Vincenti*
Marketing manager: *Dan DePasquale*

© 2004 by Pearson Education, Inc.
Publishing as Prentice Hall Professional Technical Reference
One Lake Street
Upper Saddle River, New Jersey 07458

Prentice Hall books are widely used by corporations and government agencies for training, marketing, and resale.

The publisher offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales. For more information, please contact:
U.S. Corporate and Government Sales
1-800-382-3419
corpsales@pearsontechgroup.com

For sales outside the U.S., please contact:
International Sales
1-317-581-3793
international@pearsontechgroup.com

Product and company names mentioned herein are the trademarks or registered trademarks of their respective owners.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America
First printing.

ISBN 0-13-046099-0

Pearson Education LTD.
Pearson Education Australia PTY, Limited
Pearson Education Singapore, Pte. Ltd.
Pearson Education North Asia Ltd.
Pearson Education Canada, Ltd.
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan
Pearson Education Malaysia, Pte. Ltd.

T.I.C.
LEARNING CENTRE
Book no. 32655150
Subject no. 004.6 Gan
MILLENNIUM POINT

CONTENTS

Preface xiii

Part 1 Multimedia Applications and Quality of Service (QoS) 1

Chapter 1 Multimedia Applications 3

1.1 Applications 3
1.1.1 Evolution 3
1.1.2 Video and Audio Streaming 5
1.1.3 Peer-to-Peer Computing 5
1.1.4 Digital Money Transactions 6
1.1.5 Entertainment 7
1.1.6 File and Picture Sharing 7
1.1.7 Email and Multimedia Messaging 8
1.1.8 Wireless Gaming 8
1.1.9 Voice and Telephony 9
1.1.10 Location-Based Services 9
1.1.11 Telemedicine 12
1.1.12 Business Applications 12
1.1.13 Personal Area Device Connectivity 12
1.1.14 Telematics (Automobiles) 13
1.2 Main Protocols 15

1.2.1 Short Message Service (SMS) 15
1.2.2 Enhanced Messaging Service (EMS) 16
1.2.3 Multimedia Messaging Service (MMS) 17
1.2.4 Wireless Application Protocol (WAP) 18
Chapter 2 Quality of Service Fundamentals 21
2.1 Introduction 21
2.2 QoS Parameters 24
2.2.1 Throughput 24
2.2.2 Delay 26
2.2.3 Delay Variation 27
2.2.4 Loss or Error Rate 28
2.3 Multimedia Application Requirements 29
2.3.1 Interactive and Noninteractive Applications 29
2.3.2 Tolerance and Intolerance 30
2.3.3 Adaptive and Nonadaptive Characteristics 30
2.3.4 Application Criticality 31
2.3.5 Representation of Application Requirements 31
2.3.6 Examples of Application Requirements 32
2.4 QoS Services 35
2.4.1 Quantitative (Guaranteed) Services 35
2.4.2 Qualitative (Differentiated) Services 36
2.4.3 Best Effort Services 36
2.4.4 Per-Flow QoS Services 36
2.4.5 Per-Class QoS Services 36
2.5 Realization of QoS Services 36
Chapter 3 QoS Mechanisms 39
3.1 Introduction 39
3.2 Classification 39
3.2.1 Data Link Layer Classification 41
3.2.2 Network Layer Classification 42
3.2.3 Transport Layer Classification (5-tuplet IP Header) 43
3.2.4 Application or User Classification 44
3.3 Channel Access Mechanism 44
3.3.1 Collision-Based Channel Access 44
3.3.2 Collision-Free Channel Access 45
3.4 Packet Scheduling Mechanisms 46
3.4.1 First In First Out (FIFO) 47

3.4.2 Strict Priority 48
3.4.3 Weight Fair Queue (WFQ) 49
3.5 Traffic Policing Mechanism 50
3.5.1 Leaky Bucket 50
3.5.2 Token Bucket 51
3.6 Resource Reservation Signaling Mechanisms 53
3.7 Admission Control 54
3.8 QoS Architecture 55
3.8.1 QoS Architecture for Infrastructure Wireless Networks 55
3.8.2 QoS Architecture for Ad Hoc Wireless Networks 56

Part 2 Wireless Local Area Networks 59

Chapter 4 IEEE 802.11 61
4.1 IEEE 802.11 61
4.1.1 Architecture 64
4.1.2 Physical Layer 68
4.1.3 Media Access Control (MAC) 69
4.1.4 Physical Layer Convergence Protocol (PLCP) 72
4.1.5 QoS Support 73
4.2 IEEE 802.11e (QoS Extension) 75
4.2.1 Traffic Classification 76
4.2.2 Channel Access and Packet Scheduling 78
4.2.3 QoS Signaling 82

Chapter 5 HiperLAN 85
5.1 Introduction 85
5.2 Architecture 86
5.2.1 Network Topology 86
5.2.2 Protocol Stack 87
5.3 Physical Layer 88
5.4 Data Link Control (DLC) Layer 90
5.4.1 Radio Link Control (RLC) 91
5.4.2 Error Control (EC) 91
5.4.3 Media Access Control (MAC) 92
5.4.4 Logical Channel 95
5.4.5 Transport Channels 99
5.4.6 Resource Request and Resource Grant 100

5.5 Convergence Layer 102
5.6 QoS Support 102
5.6.1 Classification 102
5.6.2 Channel Access 102
5.6.3 Packet Scheduling 103

Chapter 6 HomeRF 105

6.1 Introduction 105
6.2 Architecture 106
6.2.1 Network Topology 106
6.2.2 Protocol Stack 108
6.3 Physical Layer 109
6.4 Media Access Control (MAC) 109
6.4.1 Channel Access 109
6.4.2 Contention-based CSMA/CA with Priority Access and Time Reservation 110
6.4.3 Reserved Time TDMA with Retransmission Option 111
6.5 QoS Support 113
6.5.1 Isochronous Data Service 113
6.5.2 Priority Asynchronous Data Service 114
6.5.3 Asynchronous Data Service 115

Part 3 Wireless Metropolitan Area Networks 117

Chapter 7 IEEE 802.16 119

7.1 Introduction 119
7.2 IEEE 802.16.1 120
7.2.1 Architecture 120
7.3 Physical Layer 121
7.3.1 Adaptive Data Burst Profiles 124
7.4 Media Access Control (MAC) 124
7.4.1 Channel Access 124
7.4.2 Bandwidth Request Mechanisms and Bandwidth Allocation 128
7.4.3 Service Flow Scheduling Service 130
7.4.4 Network Entry and Initialization 132
7.5 QoS Support 133
7.5.1 QoS Provision 133
7.5.2 QoS Mechanisms 134

7.6 IEEE 802.16a 137
7.6.1 IEEE 802.16a MAC Discussions 139

Part 4 Wireless Personal Area Networks 141

Chapter 8 Bluetooth 143
8.1 Introduction 143
8.2 Architecture 144
8.2.1 Network Topology 144
8.2.2 Protocol Stack 145
8.3 Physical Layer 146
8.4 Bluetooth Baseband 147
8.4.1 Channel Definition and Channel Access 147
8.4.2 Link Control 152
8.5 Link Manager (LM) 158
8.6 Host Control Interface (HCI) 158
8.7 Logical Link Control and Adaptation Protocol (L2CAP) 159
8.8 Higher Bluetooth Layers 160
8.8.1 RFCOMM 160
8.8.2 Service Discovery Protocol (SDP) 160
8.8.3 Audio and Telephony Control (Tel Ctrl) 160
8.9 Profiles 160
8.10 QoS Support 161
8.10.1 Classification 161
8.10.2 Channel Access 162
8.10.3 Packet Scheduling 162
8.10.4 QoS Signaling 163

Chapter 9 IEEE 802.15 165
9.1 IEEE 802.15.1 165
9.2 IEEE 802.15.3 165
9.2.1 IEEE 802.15.3 Architecture 166
9.3 IEEE 802.15.3 Physical Layer 169
9.4 IEEE 802.15.3 Media Access Control 170
9.4.1 Channel Definition and Channel Access 171
9.4.2 Starting a Piconet 174
9.5 IEEE 802.15.3 QoS Support 176
9.5.1 QoS in Contention Access Period 176

9.5.2 QoS in Contention-Free Period 176
9.6 IEEE 802.15.4 179
9.6.1 IEEE 802.15.4 Architecture 180
9.7 IEEE 802.15.4 Physical Layer 182
9.8 IEEE 802.15.4 Media Access Control 183
9.9 IEEE 802.15.4 QoS Support 185

Part 5 2.5G and 3G Networks 189

Chapter 10 GPRS 191
10.1 Introduction 191
10.2 GPRS (Rel-5) Architecture 194
10.2.1 Radio Interface Protocol Architecture 197
10.2.2 GPRS (Rel-5) Protocol Architecture 202
10.3 Physical Channel 204
10.3.1 Hyperframes, Superframes, Multiframes, TDMA Frames and Time Slots 206
10.3.2 Packet Data Channel (PDCH) 206
10.3.3 Radio Block 206
10.4 Logical, Control, and Traffic Channels 209
10.4.1 Logical Channels 209
10.4.2 Control Channels 209
10.4.3 Packet Traffic Channels 211
10.5 Media Access Control (MAC) and Radio Link Control (RLC) 211
10.5.1 TBF Establishment 211
10.5.2 Channel Access and Resource Allocation 214
10.5.3 Radio Link Control (RLC) 215
10.5.4 MAC States 217
10.6 Radio Resource Control (RRC) and Radio Resource (RR) 218
10.6.1 RRC Modes of Operation 219
10.6.2 RR Modes of Operation 220
10.6.3 Transition between Modes 221
10.7 QoS Support 222
10.7.1 QoS Profile in GPRS Bearer Service 224
10.7.2 QoS Mechanisms in the Radio Bearer Service 228

Chapter 11 UMTS 231
11.1 Introduction 231
11.2 UMTS Architecture 234
11.3 Physical Layer 236
11.3.1 Transport and Physical Channels 241
11.3.2 Dedicated Data Transmission 248
11.3.3 Random Access Uplink Data Transmission 249
11.4 Media Access Control (MAC) 253
11.4.1 Logical Channels 253
11.4.2 MAC Architecture 254
11.4.3 Examples of MAC Data Transmission 258
11.5 Data Link Layer Protocols (RLC, PDCP, and BMC) 259
11.5.1 Radio Link Control (RLC) 259
11.5.2 Packet Data Convergence Protocol (PDCP) 259
11.5.3 Broadcast/Multicast Control (BMC) 260
11.6 Radio Resource Control (RRC) 260
11.6.1 RRC States 261
11.6.2 RRC Procedures 262
11.7 QoS Support 264
11.7.1 Control Plane QoS Management Functions 266
11.7.2 User Plane QoS Management Functions 267
11.7.3 QoS Classes 267
11.7.4 QoS Mechanisms in the Radio Bearer Service 270
Chapter 12 cdma2000 275
12.1 Introduction 275
12.2 cdma2000 Architecture 276
12.2.1 cdma2000 Air Interface Protocol Architecture 278
12.3 Physical Layer 280
12.3.1 Physical Channels 283
12.4 Media Access Control (MAC) 286
12.4.1 Logical Channels 288
12.4.2 Mux and QoS Sublayers 289
12.5 Link Access Control (LAC) 289
12.6 QoS Support 289
12.6.1 Bandwidth Allocation 290
12.6.2 Packet Scheduling 290

Chapter 13 Satellite Communication 291
13.1 Introduction 291
13.1.1 DVB Return Channel System (DVB-RCS) 292
13.2 Architecture 294
13.3 Forward Link 296
13.3.1 MPEG-2 Transport Stream 296
13.3.2 Data Piping 297
13.3.3 Data Streaming 300
13.3.4 Multiprotocol Encapsulation 301
13.3.5 Data Carousel 302
13.4 Return Link 302
13.4.1 Return Link Channel Access 303
13.5 Quality of Service Support 305
13.5.1 Classification 305
13.5.2 Capacity Request Signaling and Channel Access 306
Appendix Acronyms and Abbreviations 307
Part 1: Multimedia Application and Quality of Service (QoS) 307
Part 2: Wireless Local Area Networks 308
Part 3: Wireless Metropolitan Area Networks 310
Part 4: Wireless Personal Area Networks 311
Part 5: 2.5G and 3G Networks 312
GPRS, UMTS, cdma2000 312
Satellite Communication 316
Bibliography 319
Introduction 319
Part 1: Multimedia Applications and Quality of Service (QoS) 319
Part 2: Wireless Local Area Networks 320
Part 3: Wireless Metropolitan Area Networks 321
Part 4: Wireless Personal Area Networks 321
Part 5: 2.5G and 3G Networks 321
GPRS 321
UMTS 322
cdma2000 323
Satellite Communication 323
Index 325

PREFACE

The introduction of wireless communication is dramatically changing our lives. The ability to communicate anytime, anywhere increases our quality of lives and improves our business productivity. The recent technological developments that allow us to execute bandwidth-hungry multimedia applications over the wireless media add new dimensions to our ability to communicate. This opens an array of exciting opportunities in business, residential, healthcare, education, leisure, and many other areas. Wireless videoconferencing will connect us with business partners and family members. Remote video medical consultation will enhance care in rural areas and at the accident scene. Interactive games that include video and graphics with partners over the globe will add new dimensions not only to our leisure opportunities but also to provisioning of an effective remote learning environment.

Such opportunities are possible due to the recent technology developments in 1) user device miniaturization, which enables adequate computation power and display in small mobile handheld devices, and 2) provisioning of significantly broader wireless links for carrying multimedia traffic. Such broader links have been introduced in the wide-ranging wireless networks including very short-range personal wireless networks, short-range local area networks, and longer range metropolitan, cellular, and satellite networks. Because of nature of the wireless environment, these links are shared among many users executing multiple applications, each requiring different levels of quality of service (QoS) support. Therefore, each wireless network needs to incorporate *bandwidth mediation policies* that enable QoS support to the different multimedia applications. It is interesting to note that the wireless networks' standards do not provide the algorithms required for such bandwidth mediation policies. Such policies are the motivation for writing this book.

Wireless networks are described in numerous public domain documents produced by standard organizations in the U.S., Europe, Asia, and other continents whose members include hun-

dreds of participants representing companies all over the world. These organizations include IEEE (Institute of Electrical and Electronics Engineers), ETSI (the European Telecommunications Standards Institute), and ITU (International Telecommunication Union). The fact that these standards are produced by such an impressive collaboration of participants and are public domain is the basis for making these standards a true, easy means for global communication. Next generations of wireless networks include Wireless Personal Area Networks (WPANs), Wireless Local Area Networks (WLANs), Wireless Metropolitan Area Networks (WMANs), and cellular and satellite networks (see Figure 1).

The provision of a quality experience for the end-users requires end-to-end QoS support in terms of bandwidth, delay, and delay jitter. In order to provide end-to-end QoS support, we need to provide QoS in the wide area network (the Internet) as well as in the wireless extensions (WPAN, WLAN, WMAN, and cellular and satellite networks).

This book is a comprehensive guide to understanding multimedia wireless networks. The book addresses the QoS problems and solutions, discussing the architecture, applications, and implementation of wireless networks, including a number of standards and proposed standards. In this book we introduce the basic QoS support mechanisms and the standards and standardization efforts in the aforementioned array of network technologies. We focus on describing the standards' signaling mechanisms that need to be incorporated in the bandwidth mediation policies developed by network designers. We hope that the reader will realize that the development of algorithms that provide QoS support for different multimedia applications is a very complex task. We also hope that the reader will understand the available signaling mechanisms for each one of the wireless standards. Using the knowledge base provided in this book the network developers can take the necessary steps into the development and implementation of QoS mechanisms within their target wireless networks.

This book is not intended to replace the standards' documents for any design purposes. Its sole intention is to help and simplify the introduction of such standards and to provide an easy tool for making an initial comparative analysis between the different approaches presented by the standards in supporting multimedia traffic. Since the standards contain hundreds and thousands of pages of detailed information, using our personal judgment, we had to omit significant details, which we felt are less useful. Thus, we strongly recommend that for a detailed analysis and design purposes, the reader should refer to the official standards' documents.

To the best of our knowledge, this is the first book that covers such a broad array of wireless networks with a focus on multimedia support. We certainly hope to continue and develop the presented topics further by including material that describes how network designers and operators use the tools provided by the standards for providing users with an enjoyable multimedia experience. Hence, we will appreciate your feedback and input via email to Aura Ganz (ganz@ecs.umass.edu).

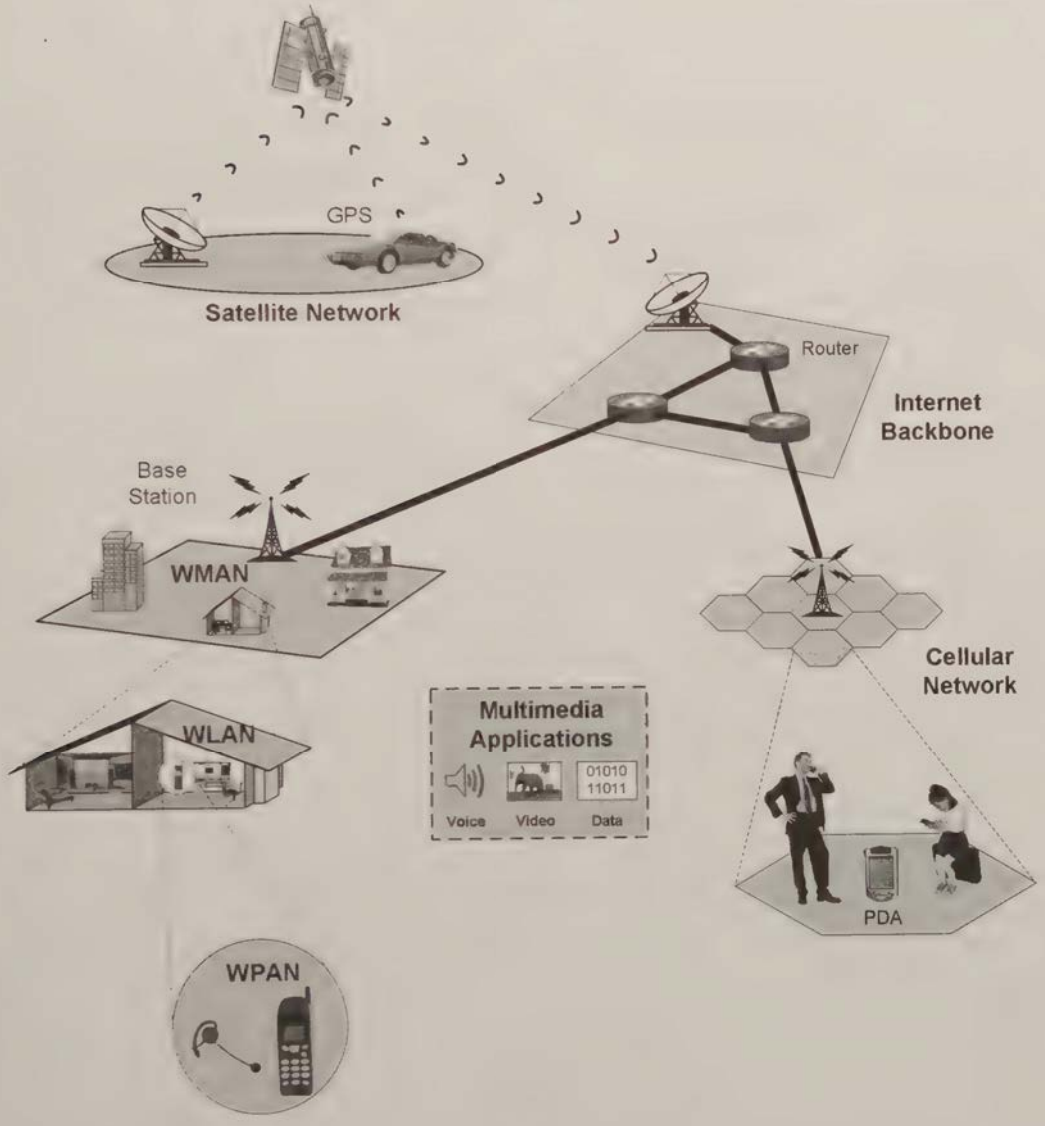


Figure 1 The Big Communication Puzzle

Target Audience

The target audience of this book includes network managers, network equipment developers, network application developers, university students, practitioners, and anyone who wants to explore QoS issues in wireless networks.

Structure of the Book

This book is organized into five primary parts.

Part 1: Multimedia Applications and Quality of Service (QoS)

In Chapter 1 we explore the nature of multimedia applications that will be delivered to the users through the wireless networks. We discuss the expectations that users have when utilizing these applications. In Chapters 2 and 3 we define QoS fundamental concepts and mechanisms and how they influence the network design. These QoS fundamentals provide the essential background required to understand the QoS support aspects of each wireless network standard in later chapters.

Part 2: Wireless Local Area Networks

In Part 2 we provide a closer look into the WLAN standards and proposed standards. Chapters 4, 5, and 6 cover IEEE 802.11, HiperLAN, and HomeRF, respectively.

Part 3: Wireless Metropolitan Area Networks

In Chapter 7 we provide a closer look into the IEEE 802.16 WirelessMAN standard.

Part 4: Wireless Personal Area Networks

In Part 4 we provide a closer look into the WPAN technologies, standards, and proposed standards. Chapters 8 and 9 describe Bluetooth and IEEE 802.15, respectively.

Part 5: 2.5G and 3G Networks

In Part 5 we provide a closer look into the 2.5G and 3G cellular networks as well as satellite network standards and proposed standards. Chapters 10, 11, and 12 cover General Packet Radio Service (GPRS), Universal Mobile Telecommunications System (UMTS), and code division multiple access (cdma2000), respectively, and Chapter 13 covers satellite networks.

PART 1

Multimedia Applications and Quality of Service (QoS)





CHAPTER 1

Multimedia Applications

1.1 Applications

1.1.1 Evolution

The landscape of Internet usage is dramatically changing because of the rapid evolution of access to wireless communication with its associated mobility. Consumers in both business and residential markets are becoming increasingly dependent on ubiquitous access. Such access is enabled by the proliferation of wireless communication. Consumers realize the benefits of many new business applications such as those in e-commerce, collaboration, supply chain, and telemedicine. They also start to appreciate the benefits for their family and personal needs such as communication with friends, entertainment, gaming, and location and safety services.

As the speed and quality provisioning of wireless communication increase, consumers' dependency on applications delivered through the wireless media will increase. They will also enjoy improved graphics as the wireless media will be able to deliver content-rich applications through their higher speed wireless networks.

This new era of ubiquitous wireless communication will provide an array of applications supported by the ability to access the Internet everywhere and anytime. Wireless communication will be available at home and at the office. For example, Microsoft installed a wireless local area network (WLAN) that lets workers connect to the corporate intranet from any spot on its 265-acre Redmond, Washington, campus. Such accessibility is becoming possible also during travel via automobile, plane, ship, and train. Wireless connectivity supports Telematics, which allows drivers to access the Internet via a screen in their cars. Whether one walks in the city, suburb, or remote desert, wireless connectivity to the Internet can be established via satellites provided by companies such as Hughes Spaceway, as they can cover any point on the continent. Other types of global networks are envisioned once high-speed 3G cellular phones become available by

companies such as AT&T (United States), Vodafone (United Kingdom), Orange (France), Omnitel (Italy), and DoCoMo (Japan). In the regional domain, Wireless Local Loop technologies will provide the basis for connection. WLANs provided by many companies such as 3Com, Linksys, and Cisco will cover local areas. Wireless Personal Area Networks (WPANs) provided by companies such as Ericsson will provide an option for coverage of personal areas in which communication needs to be established between neighboring devices.

Some of the applications that currently use wireless networks as their transmission media were previously available via wired media and some applications are new. Here is a partial list of these applications:

- Streaming video
- Streaming audio
- Collaboration
- One-way and interactive multimedia messaging
- Gaming, including interactive peer-to-peer (p2p) gaming
- Digital money transactions
- MP3 music download
- Video- and audio-supported shopping
- Long-distance learning, education
- Video and audio conferencing
- File sharing and transfer (pictures, video clips, and text)
- Feeding of real-time news and information about the weather, financial markets, sports and so on
- Geographic location services
- Safety services such as Enhanced 911 (E911)
- Gambling
- Entertainment

These applications will be delivered to consumers in many shapes and forms depending on the devices and wireless communication media available to the consumer.

The wireless network has a profound role in the effectiveness of the application delivery. When higher speed connections are available, the applications can deliver richer content, improved graphics, and more vivid colors. When the network can support Quality of Service (QoS), the applications can improve interactivity, reduce jitter, and provide continuous video and voice experience. When latency is high, applications that tailor location-based information for consumers may provide outdated information.

Likewise, the device architecture has a strong effect on the consumer experience. For example, a larger screen can accommodate larger pictures and more information. This dimension is amplified in the wireless environment, since devices tend to have much smaller screens. A more powerful central processing unit (CPU) can expedite computations associated with the applications, its graphics, and various communication protocols. More memory can enhance the graphic visualization experience. Moreover, such memory can be used for caching various multimedia contents for future view.

The potential list of applications that can benefit consumers in the wireless mobile environment is long and covers all aspects of life and business. We describe some applications in order to give an idea of the broad range of benefits that consumers can expect.

1.1.2 Video and Audio Streaming

Video and audio streaming provides the means of delivering news, entertainment, remote education, documentary, corporate speeches, fashion shows, and many more types of communication. Television may be the most well-known form of streaming video. It already feeds wireless multimedia streams into millions of dishes and antennas, connected to TVs and other devices. DirecTV (www.directv.com) and Dish Networks (www.dishnetworks.com) are two major providers of streaming video in the United States.

Streaming technologies are important, especially in the wireless world, since most users do not have access to enough connection capacity to download large multimedia files quickly. Using streaming technologies, consumers can start listening to the audio stream or view the video stream before the entire file has been received. To allow efficient streaming, the provider needs to send the data as a steady stream and the receiver needs to be able to cache excess data in a temporary buffer until used. If the data do not arrive fast enough, users will experience interruptions. There are several competing streaming technologies, such as RealAudio, RealVideo (www.real.com), Microsoft Media Player (www.microsoft.com), PacketVideo (www.packetvideo.com), and QuickTime (www.apple.com).

To reduce the amount of information transmitted, streaming video and audio data are compressed by means of technologies such as MPEG. The streaming video quality depends on the capacity of the transmission channel and its ability to support a steady stream—the better the channel quality (i.e., higher and steady data rate), the better the quality of the audio and video output.

1.1.3 Peer-to-Peer Computing

Peer-to-peer (p2p) computing is the sharing of files, memory, computation power, and other computer resources and services among devices. p2p Computing aggregates the shared network resources allowing economical execution of applications. For example, instead of purchasing additional memory for a PDA, the user can benefit from available memory on other devices.

In a p2p architecture devices can communicate directly with their peers without the need for a central server. They can support collaboration among users on different wireless and wired networks, moving data closer to the end user. This collaboration is obtained via caching mechanisms and distributed computing using the combined power of CPUs and memory of the devices that participate in the p2p network.

p2p Computing has an important role in both business and residential applications. Napster started the revolution in music distribution. Companies such as Groove, Endeavors, and eZmeeting are developing p2p business applications. Such applications support collaboration with rich content. Using p2p collaboration software, consumers can establish p2p work groups (see Figure 1.1) that are accessible only by predetermined or invited members. Through the click

of a button, workgroup members can browse the web together, share documents, edit pictures, talk to each other through a VoIP (voice over Internet protocol) feature, share calendars, or draw diagrams on a dedicated white board.

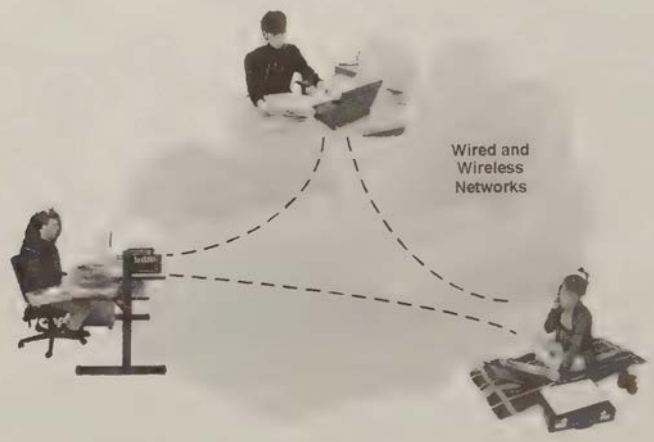


Figure 1.1 Peer-to-Peer Collaboration

1.1.4 Digital Money Transactions

Digital money transactions involving wireless devices and networks have spurred a variety of applications.

Banks will be able to support the consumer who wants to obtain cash from a traditional automated teller machine (ATM) using his or her mobile phone. To save time, the consumer is able to start the withdrawal transaction process prior to approaching the ATM. The consumer can input the account number and choose the transaction type and amount to be withdrawn. When the consumer reaches the ATM, he or she enters the security personal identification number into the mobile phone and the transaction details are transmitted to the ATM. The ATM processes the transaction and provides the amount of cash entered. Spar Nord Bank and Laan & Spar Bank in Denmark are working on such technologies.

DoCoMo and companies such as Coca-Cola introduce Cmode as part of their i-mode service. Cmode is a shopping system that uses digital transactions and that links i-mode phones to Cmode vending machines. Consumers can use their mobile phones to purchase products from Cmode-compliant machines such as vending machines which include their own computer, display screen, speakers, printers, and various sensors. To use the Cmode service, consumers need to register and deposit money into a special account maintained by DoCoMo. After the consumer makes a purchase, he or she receives a "C Ticket," which shows up on the phone's screen as a kind of bar code. When this code is passed in front of the vending machine's sensor, the machine releases the purchased product. Other communication besides the bar code transmission can be envisioned. For example, communication via a Wireless Personal Network that is established between the i-mode phone and the vending machine.

In Japan, Lawson (one of the largest chains of convenience stores with more than 7,500 shops) equipped their stores with DoCoMo iConvenience using i-mode service. With iConvenience, shoppers can make purchases through their cell phone. The shopper can enter the purchase number into the kiosk and then take the printed ticket and bar code to the cash register for payment.

1.1.5 Entertainment

One of the most prevalent applications for WLANs is entertainment. For example, home theater systems distribute movies to the monitors and speakers using wireless interconnection. The video sources (i.e., cable television, DVD player) can be in one room but the monitor can be in another room. A wireless solution makes it convenient to set up the network so people do not need to deal with wires. Digital audio files (i.e., MP3, cable television, RealAudio) have recently become popular. Instead of buying music CDs, people can purchase, download or stream digital music files from the Internet to their computers. So that the use of the entertainment is not limited to only one computer, some products provide a solution to distribute such digital music to the home stereo through a wireless audio receiver (Figure 1.2).

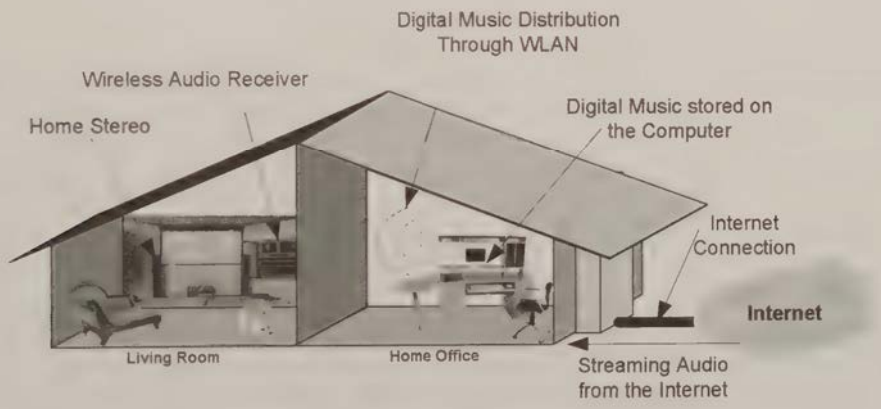


Figure 1.2 Wireless Home Entertainment

1.1.6 File and Picture Sharing

File sharing is a very common application in residential and corporate settings. Computers in the same work group or in the same domain can access and share files stored in each computer. Presently, file sharing is limited not only to a local area network, but is also extended to the Internet. There are two approaches for file sharing. One is centralized file sharing, in which there are central file servers operating as remote hard drives or online storage. A user uploads the files to a central file server. Then other users are allowed to access and download the files from the server via the Internet. For security and privacy reasons, before accessing the files, users are

required to provide some form of authentication (i.e., username or password). Companies such as Xdrive (www.xdrive.com), Yahoo! Briefcase (www.briefcase.yahoo.com), and StoragePoint (www.storagepoint.net) offer such online storage services. The other type of file sharing is distributed file sharing, in which there is no central file server. The computers connecting to the Internet can share files in a p2p fashion (as described in Section 1.1.3). A computer can access and download files directly from another computer on the Internet. Software such as Kazaa Media Desktop (www.kazaa.com) and Morpheus (www.morpheus.com) enable this kind of service.

Digital pictures generated from the digital camera or picture scanner also stimulate growth of picture-sharing services. People want to share their digital pictures with friends and family. One simple solution is to append the picture as an email attachment. However, such email files become very large, which leads to long download times, especially via modem. Companies such as Clubphoto (www.clubphoto.com), ofoto (www.ofoto.com), and Yahoo! Photos (www.photos.yahoo.com) offer users the opportunity to store and share their pictures over the Internet. Some of these services also include extra features such as photo albums, slide shows, and photo printing. These services simplify the picture-sharing process, since users need to send their friends only the URLs of the remote sites.

1.1.7 Email and Multimedia Messaging

The formation of communities has been a strong desire of many consumers, especially teens. These communities are formed around enhanced messaging services that allow participants to send messages, pictures, graphics, and video and audio files to each other. Protocols such as Short Message Service (SMS), Multimedia Messaging Service (MMS), and Enhanced Messaging Service (EMS) define the architecture for various degrees of formatting, graphics, and pictures that can be exchanged among participants.

Companies such as SpotLife (www.spotlife.com) develop video communities for wireless carriers. Participants will be able to view and send personal video content to other participants. This application can also be used for the business environment. Cell phones equipped with cameras will be able to capture live video images.

1.1.8 Wireless Gaming

The wide success of hand-held gaming devices such as GameBoy and Cybiko has demonstrated the vitality of games being played on small devices in addition to large-screen televisions and personal computers. Companies such as Sega, Nintendo, and Disney have started to target mobile wireless devices as an important part of their game development strategy. Some success stories demonstrate that wireless device users have an appetite for wireless multiplayer games.

In 2001 JAMDAT Mobile (www.jamdatmobile.com) reported that 300,000 mobile device users spent more than three million air minutes in less than three months with their multiplayer game Gladiator. Gladiator is a wireless multiplayer combat game set in the Coliseum of ancient Rome. Players choose characters and weapons for their gladiators and then duel against a live opponent in real time. The more a player utilizes a particular gladiator, the more experience and

skill this gladiator obtains. This company reported that players accessed their proprietary game server from more than 20 wireless carriers around the world using more than 65 different devices. Different players who participate in a multiplayer game may connect to different wireless networks with different capabilities. Their server allowed users to overcome such incompatibilities, providing a smooth gaming experience.

There are many other games that can be played over wireless networks, such as nGame's Rat Race (www.ngame.com) which allows a team of players to "race" against one another using the phone, Disney's Atlantis: The Lost Empire in which players create teams and search for the Lost Continent by sending bubbles up from the depths to get three black or three white bubbles in a row horizontally or vertically, and Disney/Pixar's Monsters Inc. in which players catch screams to achieve status as a Monsters Inc. professional kid scarer. Realizing the opportunity, Ericsson, Motorola, Nokia, Siemens, and other companies founded the Mobile Games Interoperability Forum (MGIF) (www.mgif.org). The forum's goal is to define specifications for mobile games developers so that their games can be played over various mobile devices and networks.

1.1.9 Voice and Telephony

Wireless networks that provide voice and telephony services include cellular phones, cordless phones and VoIP.

In 2002 the number of cellular phone subscribers reached one billion worldwide. The cellular phone offers not only voice service but also data services in the current 2.5G system and the upcoming 3G system. These systems allow users to surf the net or check email using their cellular phones.

VoIP provides voice or telephone service over a data network (e.g., wired LAN, IEEE 802.11 wireless LAN, and Internet). As shown in Figure 1.3, users can talk with each other via the computers connected to the data network. VoIP can be established either between computers or between computers and regular phones. An example of computer-to-computer VoIP software is MSN Messenger. Companies such as Net2Phone (www.net2phone.com) and Dialpad (www.dialpad.com) offer computer-to-phone VoIP software. An enterprise can deploy VoIP to reduce the cost of long-distance calls between branches (see Figure 1.4).

1.1.10 Location-Based Services

Location-based services are envisioned in many applications ranging from military to civilian everyday life. They include applications such as traffic reporting, restaurant recommendation, navigation, and customized ads. In such applications the information is sent to a user based on the user's location provided via the GPS (Global Positioning System).

United States military forces used it effectively in Operation Desert Storm. Supported by GPS, these forces were able to navigate through the wide, featureless desert in which the terrain looks the same for as much as the eyes can see. GPS allowed these forces to maneuver through sandstorms and during nighttime. This navigation system is also used for navigation by commercial applications. The construction of the tunnel under the English Channel was carried out simulta-

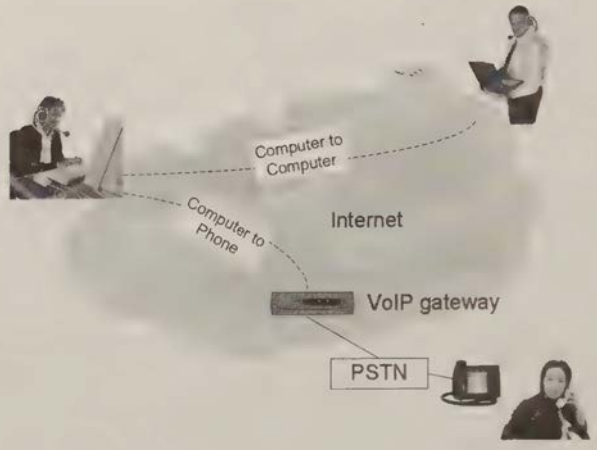


Figure 1.3 VoIP Scenario

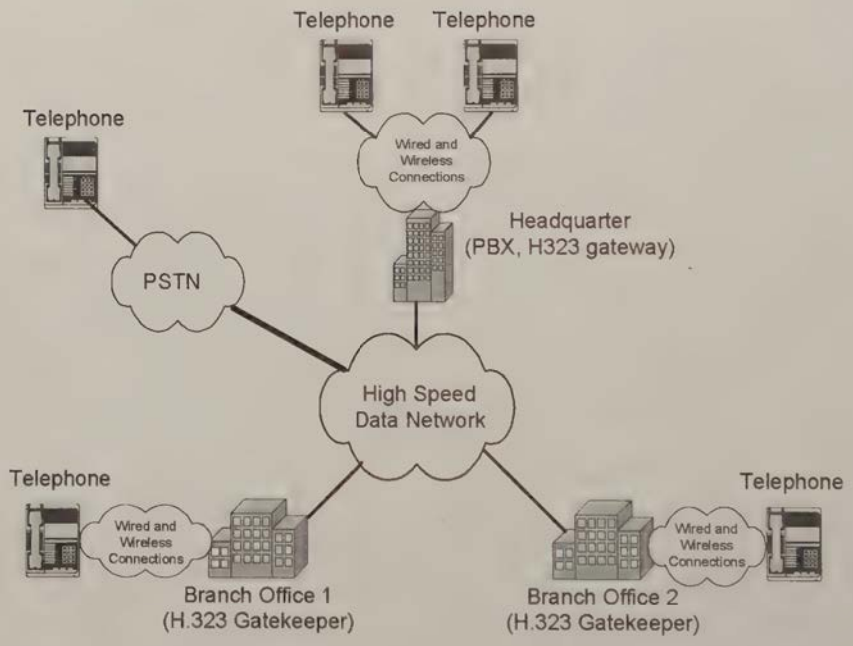


Figure 1.4 VoIP Scenario for Enterprises

neously by British and French teams digging from opposite ends of the tunnel—the English team from Dover, England, and the French team from Calais, France. These teams relied on GPS receivers outside the tunnel to validate their positions along the way and to ensure that they would meet exactly in the middle of the new tunnel. Similarly, GPS location information is being used in ships, trucks, and cars. By means of video display on the dashboard, the vehicle location data can be applied for navigation and control.

In the field of wildlife management, endangered species are being tracked by GPS receivers to monitor, study, and control population. Location information can also save human lives. One example is E911 services. They enable callers to use their cellular phones to easily place 911 emergency calls and provide the emergency services with their geographic location. In situations when the caller is traveling on an unfamiliar road, in an area with few landmarks, or at night, calling 911 on a mobile phone is probably the only choice. Using E911 services, the emergency team will have the driver's exact location, significantly shortening the time from when the rescue teams are sent until they arrive at the caller's location.

New-location based applications are invented as wireless devices equipped with GPS continue to evolve. Imagine receiving in your cell phone retail coupons based on your location. When a driver is passing a Burger King, the cell phone will beep and display an ad that suggests stopping in for one dollar off that new hamburger. The consumer can press keys to accept or refuse the offer or have an email sent for more information or future redemption. Companies such as DoCoMo (i-area service), Vindigo, and AvantGo developed such services for handheld devices. These companies report that millions of users are registered for these services.

These exciting GPS-based location services are enabled by a system of 24 satellites which are constantly orbiting the earth. They make two complete orbits every 24 hours at approximately 12,000 miles above the earth's surface. The GPS satellites continuously transmit digital radio signals that contain data on their locations and the exact time to the earth-based receivers. These satellites are equipped with atomic clocks that are very precise (less than a billionth of a second error). The receivers on earth know how long it takes for the signal to arrive. Since the signal travels at the speed of light, the receiver knows how far away the satellite is located. By using data from three satellites, the receiver computes its location in terms of longitude and latitude. By receiving a signal from one more satellite, meaning four satellites in total, the GPS receiver can also determine its altitude. The GPS system is based upon a direct line of sight between the receiver (e.g., cell phone) and the satellites. That means GPS services are not working in big cities with tall buildings or in buildings.

The FGDC (Federal Geographic Data Committee), which coordinates the development of the U.S. National Spatial Data Infrastructure, has proposed a standard for location services. This standard's objective is to increase the interoperability of location services appliances with printed maps. The committee proposes to establish a nationally consistent grid reference system as the preferred grid for National Spatial Data Infrastructure applications. This U.S. National Grid is based on universally defined coordinates and grid systems and is potentially extendable as a universal grid reference system for worldwide applications.

1.1.11 Telemedicine

Wireless communication has brought new opportunities to save lives. For example, in the remote parts of Canada, a multimedia satellite project, the Remote Communities Services Telecentre (RCST) project described in www.rcst.net, is run by a satellite operator, Telesat. RCST has been operating since 1998 linking a number of rural telecenters in Labrador and Newfoundland with others in more populated areas. These telecenters provide integrated tele-learning and telemedicine services via broadband satellite communication links. These services include high-speed Internet access, video conferencing, and digital imaging.

A physician using the RCST system in the remote Port aux Basques on the western edge of Newfoundland can, for example, transmit images of a patient's condition to a specialist in St. John's located more than 450 kilometers away. This specialist can help in diagnosing the condition, recommend treatment, or suggest a physical visit. This system saves critical time as compared to the time required for on-site diagnosis at the specialist's location, which may be especially long in difficult weather conditions.

In addition, the Integrated Emergency Medicine Network (IEMN) system has been introduced. This system provides emergency medical care for patients during ambulance transit. These special ambulances allow physicians to monitor the vital signs of patients in transit and better prepare for their arrival at the hospital (see Figure 1.5).

Similar projects have been implemented in Africa, the Philippines, and locations where the dispersed population faces difficult long-distance commutes to reach health care providers.

1.1.12 Business Applications

Business applications such as payment authorization, distance learning, multicast delivery of promotional content, and retail point-of-sale (POS) transaction processing can be handled centrally via one satellite communication provider. For example, General Motors and Ford are able to train its management and technical personnel in any office, anywhere in the world. Each student can instantly access interactive distance learning modules via DirecWay Hughes Satellites. WalMart, Shell, Texaco, and other companies use this service for their retail POS transactions.

1.1.13 Personal Area Device Connectivity

Personal area devices refer to devices carried by, placed near, or used near a person. There is a wide range of such devices:

- Consumer devices (Figure 1.6): portable CD player, digital camera, printer
- Games: game controller, joystick
- Professional devices: personal digital assistant (PDA), notebook, pager, cellular phone
- Sport training devices: health monitor, sensor, motion-tracking devices
- Hospital devices: blood pressure sensor, heart rate sensor, electrocardiogram
- Military devices: combat equipment (Figure 1.7)



Figure 1.5 Telemedicine Scenario

Typically, personal devices have proprietary cables that interconnect (e.g., earphone, printer cable). Sometimes the cable is bulky, easily damaged, and frustrating to use. Wireless technologies become alternative solutions to replace these cables.

1.1.14 Telematics (Automobiles)

New cars are designed with more and more digital devices for every aspect including safety, operation, communication, and entertainment. There are new digital devices that control every function of the car ranging from fuel injection to airbag operation. In addition, entertainment options have increased to include satellite radio, video, and gaming. Wireless communication, combined with location services provided by GPS, has enabled safety and security services such as routing instructions, emergency door unlocking, emergency response, and stolen vehicle notification (see Figure 1.8).

Mercedes-Benz offers voice-activated phones, automatic headlights, and sensors that watch for nearby obstacles. BMW offers email access as well as its iDrive system, which allows drivers hundreds of options to access the car devices with a small control knob and on-screen menus. In addition, BMW's Mayday system offers wireless communications and GPS location services for 24-hour emergency response, roadside assistance, stolen vehicle notification, and remote door unlocking.

Similar services are offered by other companies such as General Motors. Their *OnStar* System integrates on-board advanced vehicle electronics with GPS technology and wireless communications technology. The OnStar System operates 24 hours a day and links the driver and car to a 24-hour response center for safety, security, and convenience needs.

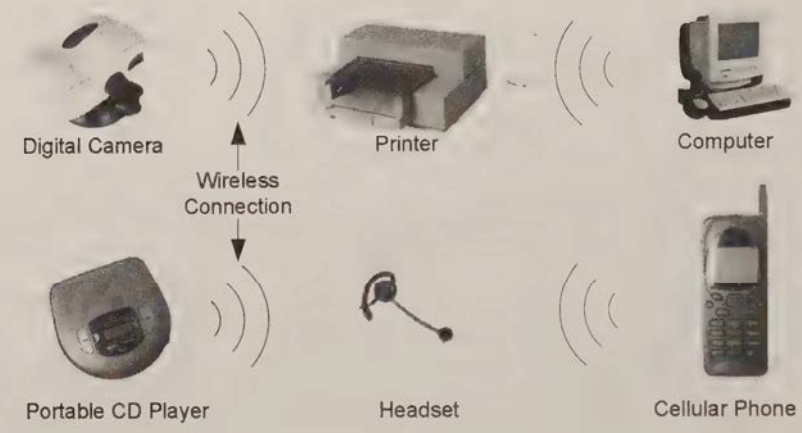


Figure 1.6 Consumer Devices



Figure 1.7 Combat Equipment

Motorola's iRadio project demonstrates wireless delivery of a wide variety of services. The iRadio converges into one package that includes picture entertainment, information, navigation, and communication. Drivers and passengers can configure their personal preferences. Its navigation system can provide real-time route planning and travel instructions, off-route notification, destination identification, traffic advisories, and local information. Radio is provided by the satellite radio channels. Users have access to news, stock quotes and other financial updates. Telephony services including email access are also supported. Internet connection can be provided through satellites or cellular network connections.

Ford, Daimler-Chrysler, General Motors, Renault, and Toyota formed the Automotive Multimedia Interface Collaboration (AMIC) forum to facilitate the development, promotion, and standardization of automotive multimedia interfaces to motor vehicle communication networks.

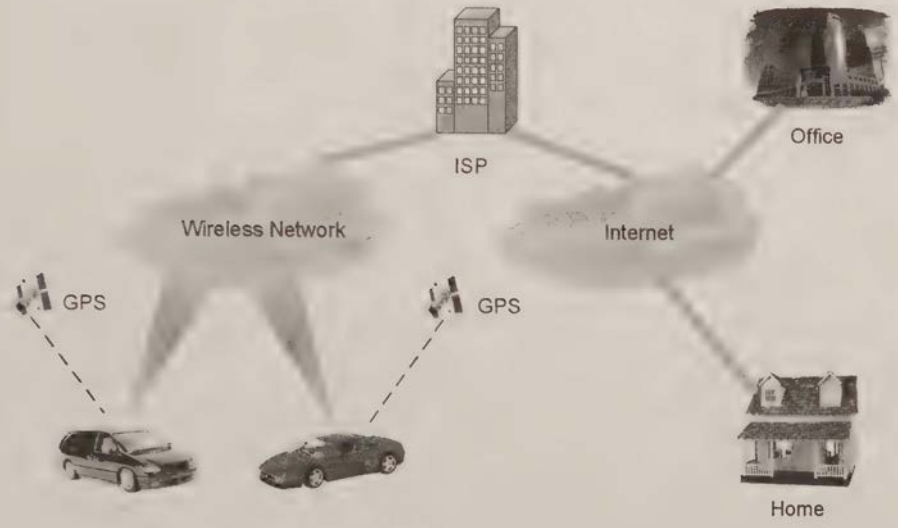


Figure 1.8 Telematics Scenario

The goal is to have an open standard that will eliminate complex and incompatible systems that can only operate on a single type of vehicle. The open standard will make it possible for electronic devices to be automatically configured and to communicate with each other. The forum considers a variety of wired and wireless technologies for the networked vehicle.

1.2 Main Protocols

1.2.1 Short Message Service (SMS)

Short Message Service (SMS) is a text message service that enables the transmission of short messages between a cell phone, PDA, PC, or any device with an Internet protocol (IP) address. Since this service was standardized in the mid 1990s, it has recorded tens of billions of messages and is being used worldwide. SMS was introduced in the Global System for Mobile Communications (GSM) and later adopted by other mobile communication systems such as Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA).

The service supports messages that are no more than 140 to 160 alphanumeric characters in length, which is typically around 30 to 40 words in Latin and 70 characters for non-Latin alphabets like Chinese and Arabic. Initial applications of SMS were based on replacing alphanumeric pagers by supporting two-way messaging services, primarily voice mail.

SMS is used for many applications. For example, GameWorld Technologies' game entitled "Women Are Smarter?" is a mobile trivia game with a gender twist. Players respond to trivia questions that include the option to make double-or-nothing wagers on which gender answers the question correctly more often. Data regarding question and regarding which gender answers correctly more frequently are updated in real-time.

Additional services were added such as electronic mail, fax, notification service in which users are notified when predetermined events occur (e.g., receipt of email, scheduled appointments), integration of email into the SMS service, paging and its integration into the SMS service, interactive banking, and information update services (e.g., financial information, weather, news, directory assistance).

The SMS service on the SMS-enabled phone is always on—the phone is able to receive or send a short message at any time, even if a voice conversation is in progress. This is possible since the SMS messages are delivered to and from the cell phones or wireless devices over the system's control channel and not over the voice channel. This helps in preserving the voice quality while utilizing the control channel. However, the control channel capacity imposes a limit on the length of the messages since the control channel is used for control services such as phone location and call management. As the numbers of SMS messages have increased, concerns about potential service problems caused by overdemand have been mentioned.

SMS guarantees delivery of the short message. Temporary failures are identified, and the short message is stored in the network until the destination device becomes available. The message is sent to the nearest Short Message Service Center (SMSC), which either delivers it to the addressed mobile device or forwards it to the next SMSC. The messages are stored and forwarded in the SMSCs until they are received by the addressees. In the delivery process, the SMSC first sends an SMS request to the Home Location Register (HLR) to locate the addressed mobile phone. The HLR is a database used for storing and managing subscriptions and service profiles. If the addressed mobile phone is found, the HLR provides the routing information. If the HLR cannot locate the addressed phone, then the SMSC stores the message for a limited time and tries to deliver the message again when the addressee connects to the network. If the addressee is found, the message is delivered and the SMSC receives verification that the message has been received by the addressee. The sender is informed that the SMS message has been received by the addressee.

1.2.2 Enhanced Messaging Service (EMS)

Enhanced Messaging Service (EMS) evolved from the popular SMS. Service providers started to offer EMS in early 2000. EMS allows much richer content options than text-based SMS messages and can include pictures, melodies, sound marks, graphic, animations, fonts, and formatted text. For example, a person could send this text to his date with an SMS message "I love you." Using EMS, he could add a short animation of flowers along with a love melody.

EMS uses the store-and-forward mechanism provided by the SMSCs. Similar to SMS, EMS uses the control channels. EMS is enabled by SMS concatenation, linking several short messages together.

EMS supports basic and extended pictures. Basic pictures are black and white small (16x16 pixels) pictures or large (32x32 pixels) pictures. Extended pictures can be, however, black and white, grayscale, or color. Extended pictures can contain 255x255 pixels and can be transmitted in a compressed format.

EMS also supports sound. There are a few predefined sounds, including low and high chimes and chords, Claps, TaDa, and drum sounds, as well as Notify, User-Defined, and Extended Sounds. Predefined sounds are not transmitted over the air. Only a reference to the sound is included in the EMS. There are ten different sounds that can be added in the message, and when the sound reference is being displayed, the referenced sound will be played. The sender can also download melodies from various web sources. These sounds have to be formatted according to the iMelody standard. These melodies can take up to 128 bytes.

Animations in EMS provide users with a much stronger alternative for expression than plain pictures. EMS supports predefined animations that reflect happiness, sadness, flirtatiousness, gladness, skepticism, and grief. User-defined and extended animations are supported as well. Similar to sounds and pictures, EMS predefined animations are not sent as animation over the air. Only a reference to them is included in the EMS message. When the message is received by the addressee the referenced animation is displayed in a manner that is specified by the manufacturer. User-defined animations consist of four pictures. There are two different animation sizes: small (8x8 pixels) animations and large (16x16 pixels) animations. These animations are sent over the air interface. Extended animations may be black and white, grayscale, or color. The maximum size of a single animated frame is 255x255 pixels. The repetition of these animations may be controlled by the sender and can be transmitted in a compressed form.

An EMS message can be sent to a mobile phone or other wireless device, even if it does not support EMS, because all the EMS components (i.e., text formatting, pictures, animations, and sounds) are located in the message header. The EMS contents included in the header will be ignored by the receiving mobile phone if it does not support EMS, and only the text message will be displayed.

1.2.3 Multimedia Messaging Service (MMS)

Multimedia Messaging Service (MMS) technology is the ultimate messaging application, allowing users to create messages that include any combination of text, graphics, photographic images, speech, and audio or video clips. MMS supports standard image formats such as JPEG and GIF, video formats such as MPEG 4, and audio formats such as MP3 and MIDI. Multimedia messaging depends on the high transmission speeds that will be available via 3G technologies.

MMS is expected to be the future multimedia messaging technology. As compared with the SMS size limitation of 160 bytes, MMS will have no limitation even if some initial implementations may pose restrictions on the message size. Rather than sending a simple message such as "I am late," the user will be able to send a message that better explains the reasons of being late. This unlimited message size will allow users to express rich content in their messages for ultimate expression of ideas, personality, and feelings. Being able to send pictures allows users to share experiences with friends, family members, and business partners. In business applications, users can capture relevant pictures and send them to colleagues or to their home office or they can store them locally for future retrieval.

MMS built-in presentation layering can control timing and synchronization, allowing users to view, listen to, and read the messages simultaneously. MMS can support picture sharing.

MMS will be able to include still images such as pictures, screensavers, postcards, graphics, greeting cards, maps, and business cards. In addition, MMS will support animation, video, cartoons, and interactive video.

MMS will use the Wireless Application Protocol (WAP) as supporting technology and the high-speed 2.5G and 3G transmission technologies such as Enhanced Data rate for Global Evolution (EDGE), General Packet Radio Service (GPRS), and Universal Mobile Telecommunications System (UMTS). These high-speed connection technologies can provide users with the necessary bandwidth to send and receive rich-content multimedia messages. In SMS and EMS, the messages are transmitted over the control channel, which severely limits the transmission capacity. Instead, MMS will use the data channels used by all other voice and data applications. Consequently, MMS will be able to deliver much larger messages. MMS is also expected to use MExE (Mobile Execution Environment), which is a flexible and secure application environment for 2.5G and 3G mobile devices. MExE includes a variety of current technologies such as WAP and Java. Similar to SMS and EMS, MMS is a non-real-time service that routes multimedia messages to MMS servers.

MMS can include the following:

- Audio, sound, streaming audio, melodies, songs. For example, instead of sending a downloaded birthday jingle in EMS, a user can send a personalized video clip combined with the song "Happy Birthday."
- Images or pictures that are either stored or taken via a camera that is attached or part of the wireless phone or terminal. Users can take a snapshot and immediately send it to an addressee.
- Video clips. Initially the video clip may be limited to 30 seconds. This will allow users to include video clips recorded with their digital camera and to add text, labels, and appropriate audio to their messages. MMS will also support streaming video so that users can subscribe to news and entertainment services and send their own captured video as a video stream.
- Synchronized Multimedia Integration Language (SMIL, pronounced "smile") which enables simple authoring of interactive audiovisual presentations. SMIL can be used to create rich-content multimedia presentations that include streaming video and audio as well as pictures, images, text, or any other medium type. SMIL is based on an HTML-like language (standard defined by www.w3.org).

1.2.4 Wireless Application Protocol (WAP)

Wireless Application Protocol (WAP) is an open, global standard that provides a microbrowser environment optimized for wireless devices, such as phones and pagers. The goal is to easily access and interact with information over the Internet. WAP is being employed in the broad area of wireless applications described in this chapter. This standard is defined by the WAP forum whose members are wireless and Internet companies around the world to ensure interoperability

and foster growth of the wireless markets. It ensures that all WAP-based applications work across all devices, from cell phones to more powerful hand-held devices. Also it allows for WAP-based applications to use less memory, processing power, display, and key handling than legacy Internet applications. Thus the WAP standard helps in delivering the multimedia content while using less bandwidth.

The most recent standard is WAP 2.0, published in 2002. It adds support for the standard Internet communication protocols: IP, TCP, and HTTP. It also allows applications to work over all existing and coming wireless 3G technologies. It provides a rich application environment, which enables delivery of information and interactive services to digital mobile phones, pagers, PDAs, and other wireless devices. It addresses the unique characteristics of wireless devices. These devices have hardware limitations (e.g., small screens, limited battery life, and limited memory) requiring special attention to user interface design (e.g., one-finger navigation).

WAP uses the Pull Model, where the user requests content from the server. WAP 2.0 adds telephony support with WTA (Wireless Telephony Application) which enables a wide range of advanced telephony applications in addition to their legacy support of data-only functionality. These tools include call handling services, such as making calls, answering them, placing them on hold, and redirecting them.

WAP 2.0 supports a Push Model, which allows server-based applications to send or "push" the content to the devices via a Push Proxy (see Figure 1.9). Push functionality is especially important for sending news, announcements, and notifications to interested users. The content can include stock prices, location-based promotions, and traffic update alerts. Without push functionality, these types of applications would require the devices to specifically pull or request application servers for such new information. In wireless environments such pulling activities, if done persistently and frequently, would burden the limited resource network with wasteful traffic.

The External Functionality Interface (EFI) service specifies the interface between the Wireless Application Environment (WAE) and plug-in modules, which extends or enhances the capabilities of browsers or other applications. The EFI framework provides for future growth and extendibility of supported WAP devices and can be used to access external devices (e.g., smart cards, GPS devices, digital cameras).

A WAP proxy (or WAP gateway) was required in the original WAP standard (see Figure 1.10). Such proxy handles the conversion between the WAP client and the origin server equipped with legacy Internet servers that are not optimized for the limited bandwidth of the wireless channel. WAP 2.0 does not require a WAP proxy. However, use of a WAP proxy can provide several important benefits such as optimizing the communication process in the wireless network and may offer mobile service enhancements such as location, secure channels (privacy), caching, and presence-based services. A proxy can also translate between WAP and other WWW protocols, allowing WAP clients to communicate with servers that do not support WAP. A WAP proxy is necessary to offer push functionality. Proxies may be located in several locations, for example, at the wireless telephone service providers and hosting companies.

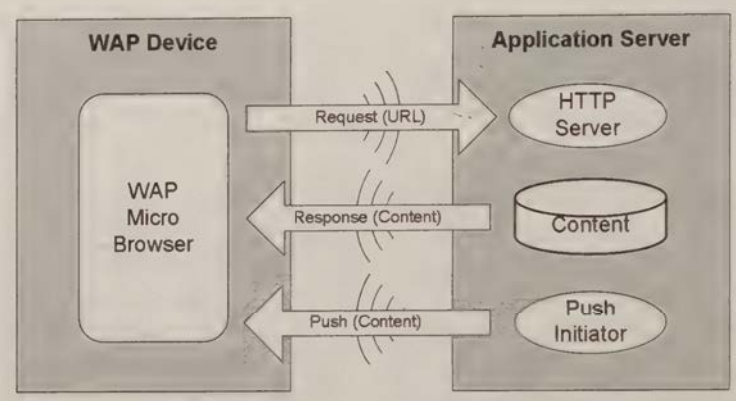


Figure 1.9 The WAP Programming Model

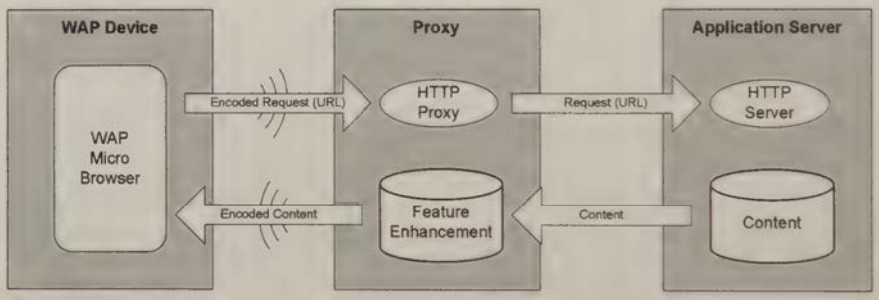


Figure 1.10 WAP Optional Proxy Model

CHAPTER 2

Quality of Service Fundamentals

2.1 Introduction

The previous chapter introduced a wide variety of multimedia applications used in wireless networks. These applications emerge to serve people's needs and sometimes create new services that attract people to deploy wireless networks in their businesses or daily lives. Multimedia applications discussed in this book mainly focus on *network applications,* i.e., applications where *hosts* and *machines* communicate through a *network.* Unlike the stand-alone multimedia applications where the multimedia contents are originated and displayed on the same machine, the network multimedia application contents that originate on a source host are transmitted through the network and displayed at the destination host. Therefore, there are a number of factors and components that affect the performance of multimedia applications such as:

- *Users:* The (human or nonhuman) ones who utilize the multimedia applications. Users' perception can influence the evaluation of the multimedia applications' performance.
- *Host machine:* The devices that operate the multimedia application (source and destination hosts). The hosts consist of a number of components such as processors, media storage systems (e.g., hard drive, CD-ROM), display devices, and operating systems.
- *Application:* The structure or the mechanisms built in the multimedia application (e.g., the codec used in video compression).
- *Network:* The network components that transport the multimedia contents between the two host machines (source and destination). Examples of hardware components include: switches, routers, network interface cards, gateways, and firewalls. The network also includes network protocols that reside at each network hardware element.

Figure 2.1 shows a simplified diagram of a communication system that includes all the components mentioned above. Hosts connect to the network through the network interface devices using various network technologies (e.g., IEEE 802.11 WLAN, Satellite, Bluetooth). The network cloud consists of multiple segments of interconnected subnetworks that establish the communication path between the hosts. All of these components require multimedia support.

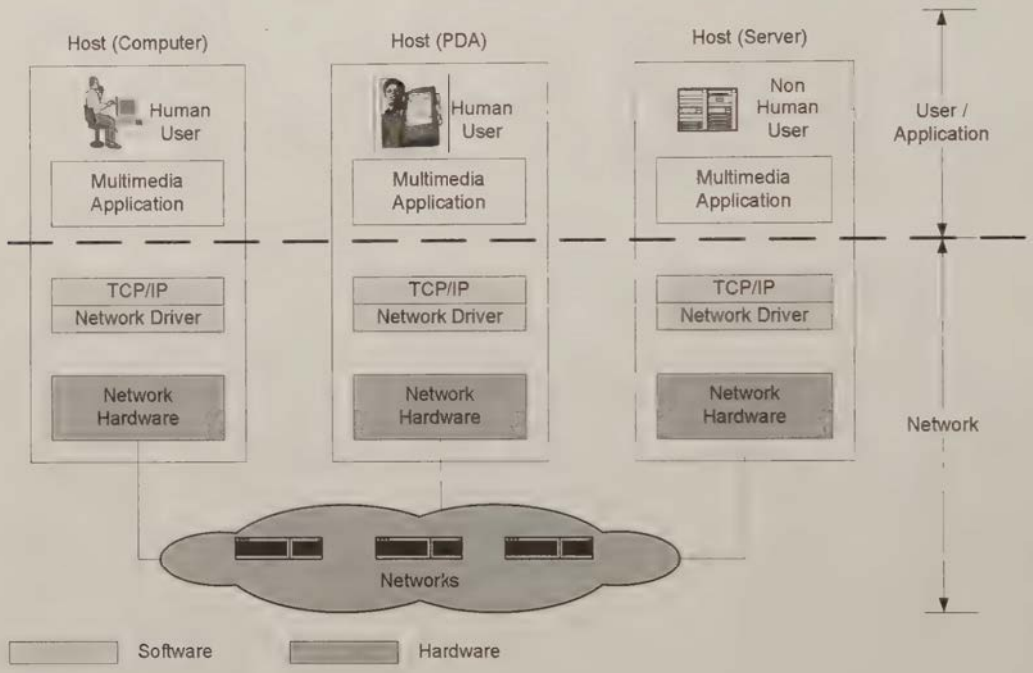


Figure 2.1 Communication System

Multimedia support issues can be presented by using the Quality of Service (QoS) term, which is an overloaded term with various meanings and perspectives. There is little consensus on the precise definition of QoS. Different people and communities perceive and interpret QoS in different ways. For example, in the networking community, QoS refers to the service quality or service level that the network offers to applications or users in terms of network QoS parameters, including latency or delay of packets traveling across the network, reliability of packet transmission, and throughput. However, in application communities, QoS generally refers to the application quality as perceived by the user—that is, the presentation quality of the video, the responsiveness of interactive voice, and the sound quality (CD-like or FM-radio-like sound) of streaming audio. Instead of providing a precise QoS definition, we present a simplified QoS model that includes two QoS perspectives: Application/User and Networks (see Figure 2.2).

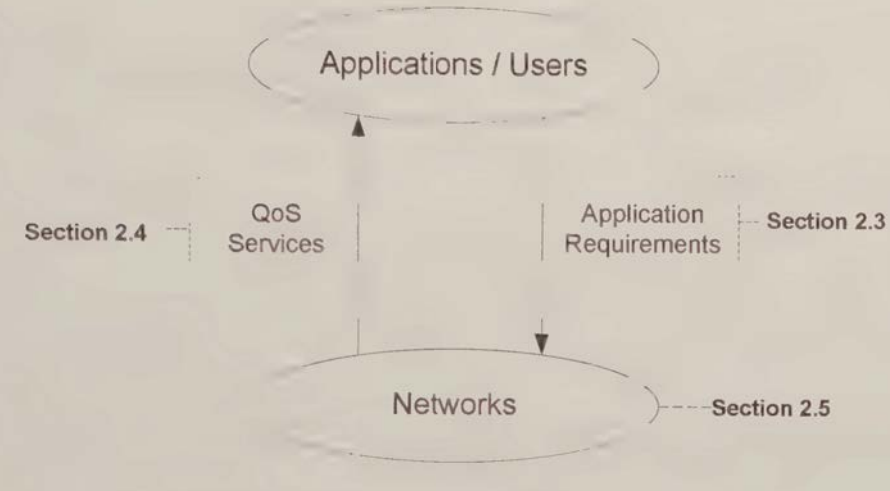


Figure 2.2 QoS Model and Corresponding Section in This Chapter

We assume that applications and users are in the same group because of their close relationship and the common way they perceive quality. In this book we often use the terms "applications" and "users" interchangeably. The applications/users expect a specific QoS in terms of response time, for example consistent perceived quality and uninterrupted service. These QoS requirements are passed to the network implicitly or explicitly and the underlying networks are responsible in part for meeting these requirements.

Users are not concerned about how the network manages its resources or what mechanisms are involved in QoS provision. However, the users are concerned about the services that networks provide which directly impact the perceived quality of the application.

From the network perspective, the networks' goal is to provide the QoS services that adequately meet the users' needs while maximizing the network resources (i.e., bandwidth) utilization. To achieve this goal, the networks analyze the application requirements, manage the network resources, and deploy various network QoS mechanisms.

Although the QoS model here looks rather simple, there are a number of subtle issues that need to be discussed. For example:

- What information is contained in the application requirements?
- Desired quality of the application which relates to users' satisfaction is subjective. How does the desired quality map to QoS parameters which are managed by networks? What criteria need to be considered in the mapping process?
- What are the QoS parameters?
- Which QoS mechanisms do the networks use to achieve diverse QoS support?
- What kind of QoS service level can networks offer?
- What is the relationship between QoS services and application requirements?

In the remaining part of this chapter we attempt to answer these issues. This chapter is organized as follows. In Section 2.2 we introduce the application and network QoS parameters. Section 2.3 discusses multimedia application requirements in terms of QoS parameters. QoS services are introduced in Section 2.4, and the realization of QoS services is described in Section 2.5.

2.2 QoS Parameters

The following QoS parameters are relevant to multimedia applications:

1. Throughput or bandwidth
2. Delay or latency
3. Delay variation (delay jitter)
4. Loss or error rate

2.2.1 Throughput

From the application perspective, throughput refers to the data rate (bits per second) generated by the application. Throughput, measured in the number of bits per second, sometimes is called *bit rate* or *bandwidth*. Bandwidth is considered to be the network resource that needs to be properly managed and allocated to applications.

The throughput required by an application depends on the application characteristics. For example, in a streaming video application, different video properties generate different throughput. A user can select the video quality by varying the following video properties:

- *Frame size:* A function of the number of pixels in each row and column and of the number of bits per pixel.
- *Frame rate:* The refreshing video frame rate (number of frames per second). Decreasing video frame rate reduces the bandwidth consumption but compromises the smoothness of the video movement.
- *Color depth:* The number of possible colors represented by a pixel. The 256-color video requires 8 bits of data per pixel; whereas, the 16-million-color video requires 24 bits of data per pixel.
- *Compression:* Reduction of the bandwidth consumption at the expense of image quality. Examples of video compression standards include MPEG1, MPEG2, and MPEG4.

In the remaining part of this subsection we will describe two characteristics of an application data traffic generation process: the data traffic generation rate (constant and variable bit rate) as well as data traffic generation burstiness. We will show how the traffic generation process determines the applications' required throughput.

2.2.1.1 Constant and Variable Bit Rate

Constant bit rate (CBR) applications generate data traffic with constant data rate. Examples of CBR applications include the following:

- Digital telephone Private Branch Exchange (PBX), which generates 64 kbps constant bit rate
- Uncompressed digital video

Most of the constant bit rate applications are delay sensitive and require constant bandwidth allocation. Allocating bandwidth below the required bandwidth causes application failure. On the other hand, allocating bandwidth above the requirement does not improve the user satisfaction, as shown in Figure 2.3.

Variable bit rate (VBR) applications generate data traffic with variable data rate. The degree of bit rate variability depends on the application. Examples of VBR applications include the following:

- Compressed video and audio
- Remote login

VBR applications require minimum bandwidth allocation in order to operate successfully. The more allocated bandwidth, the better the user-perceived quality. Bandwidth allocation beyond the maximum required bandwidth does not improve user satisfaction as shown in Figure 2.3.

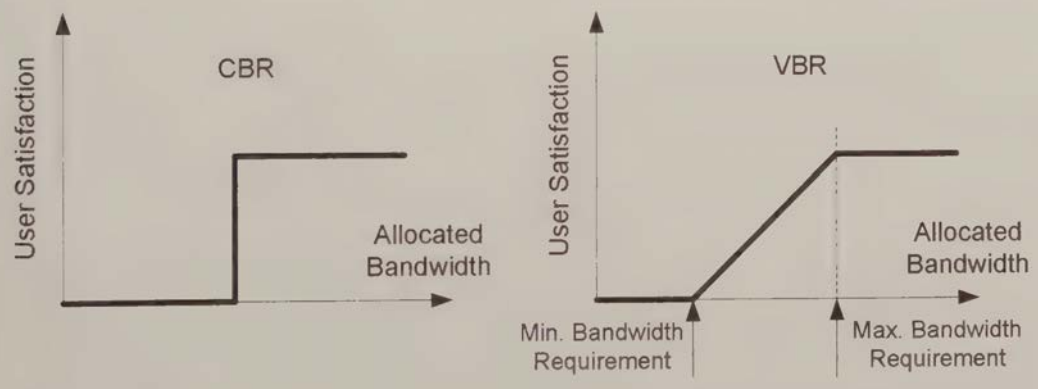


Figure 2.3 User Satisfaction as a Function of Allocated Bandwidth for CBR and VBR traffic

2.2.1.2 Burstiness

The data traffic *burstiness* measures the degree of bit rate variability of a VBR application. The burstiness is defined as the ratio between the Mean Bit Rate (MBR) and Peak Bit Rate (PBR) where:

- PBR is the maximum number of bits in a short period of time
- MBR is the average number of bits in a long period of time

2.2.2 Delay

Delay has a direct impact on users' satisfaction. Real-time applications require the delivery of information from the source to the destination within a certain period of time. Long delays may cause incidents such as data missing the playback point, which in turn reduces the video fidelity. Moreover, it can cause user frustration during interactive tasks. When the data traffic is carried across a series of components in the communication system that interconnects the source and the destination, each component introduces delay. We can categorize the main sources of delay as follows:

1. *Source-processing delay* (digitization and packetizing delay): This delay, which is introduced by the source that generates the packets, depends on the source host hardware configuration (CPU power, RAM, motherboard, etc.) and its current load (e.g., the number of applications running simultaneously and their required hardware resources).
2. *Transmission delay:* The transmission time of a packet is a function of the packet size and transmission speed.
3. *Network delay:*
   a. *Propagation delay:* The propagation delay from the source to the destination is a function of the physical distance between the source and the destination.
   b. *Protocol delay:* The delay is caused by the communication protocols executed at the different network components such as routers, gateways, and network interface cards. The delay depends on the protocols, the load of the network, and the configuration of the hardware that executes the protocol.
   c. *Output queuing delay:* The delay is caused by the time a packet spends in the outgoing link queue at a network component. For example, such delay can be incurred at an intermediate router output queue. The delay depends on the network congestion, the configuration of the hardware, and the link speed.
4. *Destination processing delay:* This delay is introduced by the processing required at the destination. For example, such delay can be incurred in the packet reconstruction process. Similar to the processing delay at the source, this delay depends on the destination host hardware configuration and load.

Figure 2.4 depicts the end-to-end delay diagram.

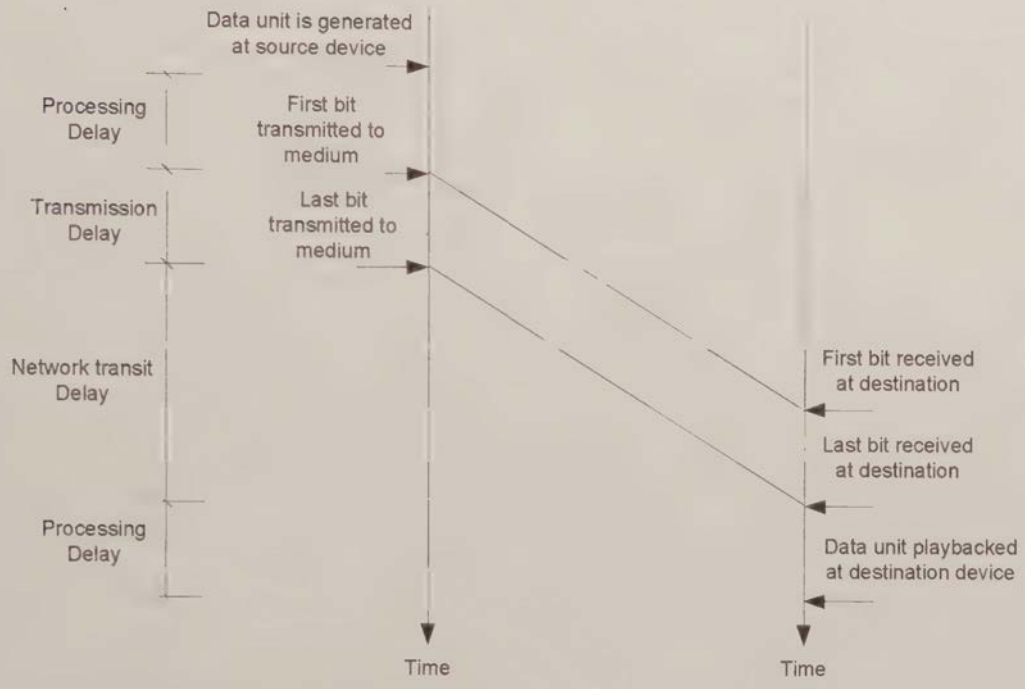


Figure 2.4 End-to-end Delay Diagram

2.2.3 Delay Variation

Delay variation is a QoS metric that refers to the variation in the delay introduced by the components along the communication path. Since each packet in the network travels through different paths, and the network conditions for each packet can be different, the end-to-end delay varies. For data generated at constant rate, the delay jitter distorts the time synchronization of the original traffic. As shown in Figure 2.5 the packets travel through the network and experience different end-to-end delays, reaching the destination with timing distortions (incomplete or delayed signal) relative to the original traffic. There are several techniques to cope with delay jitter at the receiver end. For example, in technique A (Figure 2.5A), the receiver playbacks the signal as soon as the packets arrive. The playback point is changed from the original timing reference. This introduces distortion in the playback signal. In technique B (Figure 2.5B), the receiver playbacks the signal based on the original timing reference. The late packets that miss the playback point will be ignored. This also introduces distortion. In technique C (Figure 2.5C) a de-jittered buffer is used. All packets will be stored in the buffer and held for some time (offset delay) before they are retrieved by the receiver with the original timing reference. The fidelity of the signal will be maintained as long as there are packets available in the buffer. Large delay jitter requires large buffer space to hold the packets and smooth out the jitter. A large buffer also

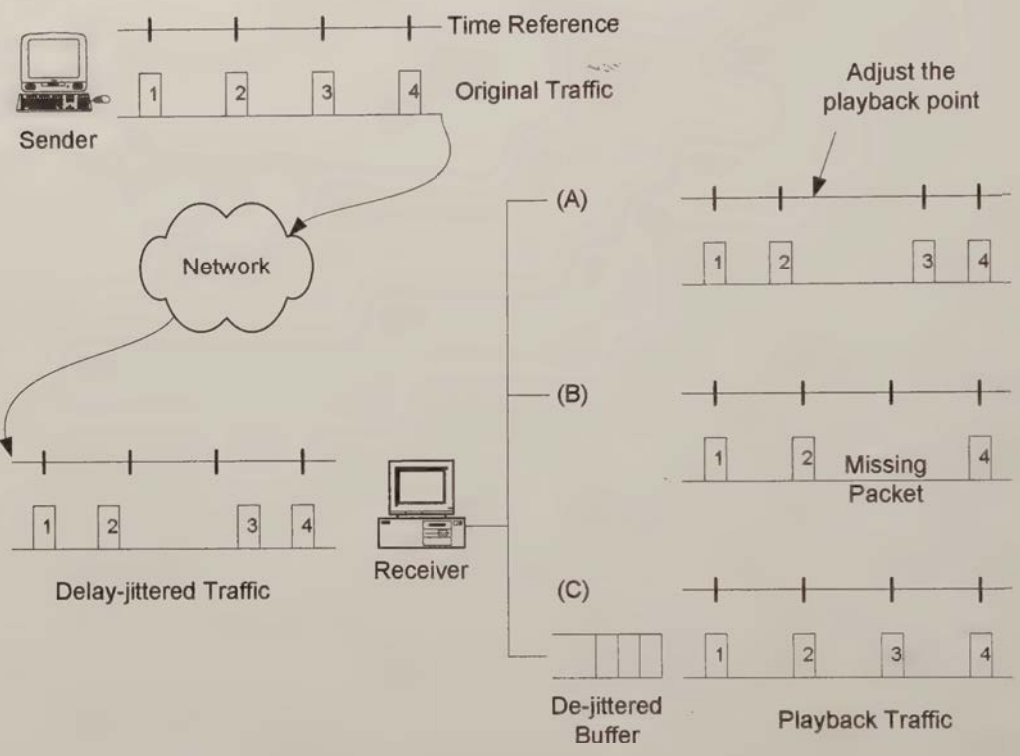


Figure 2.5 Delay Jitter and Its Solutions

introduces large delays, which will be eventually constrained by the application delay requirement. In summary, there is a tradeoff between the following three factors: de-jittered buffer space, delay requirement, and fidelity of the playback signal.

2.2.4 Loss or Error Rate

Packet loss directly affects the perceived quality of the application. It compromises the integrity of the data or disrupts the service. At the network level, packet loss can be caused by network congestion, which results in dropped packets. Another cause of loss is caused by bit errors that occur due to a noisy communication channel. Such loss will most likely occur in a wireless channel. There are several techniques to recover from packet loss or error such as packet retransmission, error correction at the physical layer, or codec at the application layer that can compensate or conceal the loss.

2.3 Multimedia Application Requirements

Different multimedia applications have different QoS requirements expressed in terms of the following QoS parameters as described in the previous section: throughput, delay, delay variation, and loss. In many cases, users can determine the application's QoS requirements by investigating the factors that influence the application quality (i.e., the task characteristics, user characteristics). For example, from experimentation they conclude that for acceptable quality, the one-way delay requirements of interactive voice should be less than 250 ms. This delay value includes the delay introduced from all components in the communication path such as source delay, transmission delay, network delay, and destination delay.

In this book, we focus on the applications' requirements for multimedia support in wireless networks. Due to advances in the source and destination available processing power and the very high-speed optical infrastructure, we expect delay and bandwidth bottlenecks to occur in the wireless network due to their limited available bandwidth.

Before presenting detailed QoS requirements for each application, we would like to provide an overview of the factors that influence the application requirements. Some of the factors are as follows:

- *Application interactivity level:* Interactive and noninteractive applications
- *User/Application characteristics*: Delay tolerance and intolerance, adaptive and nonadaptive characteristics
- *Application criticality:* Mission-critical and non-mission-critical applications

We will describe below each one of these factors.

2.3.1 Interactive and Noninteractive Applications

An interactive application involves some form of interaction (action-reaction, request-response or exchange of information) between two parties (people-to-people, people-to-machine or machine-to-machine). Examples of interactive applications include the following:

- *People-to-people application:* IP telephony, interactive voice/video, videoconferencing
- *People-to-machine application:* Video-on-demand (VOD), streaming audio/video, virtual reality
- *Machine-to-machine application:* Automatic machine control

The elapsed time between interactions is essential to the success of an interactive application. The degree of interactivity determines the level or stringency of the delay requirement. For example, interactive voice applications, which involve human interaction (conversation) in real time, have strict delay requirements (in the order of milliseconds). Streaming (playback) video applications involve less interaction, (i.e., interaction mostly occurs during start, stop, forward, or reverse action on the video) and do not require real-time response. Therefore streaming appli-

cations have more relaxed delay requirements (in the order of seconds). Often the applications' delay tolerance is determined by the users' delay tolerance (i.e., higher delay tolerance leads to more relaxed delay requirements).

Delay jitter is also related to QoS support for interactive tasks. As discussed above, the delay jitter can be corrected by de-jittering buffer techniques. However, the buffer introduces delay in the original signal, which eventually affects the interactivity of the task. In general, an application with strict delay requirements also has strict delay jitter requirements.

2.3.2 Tolerance and Intolerance

Tolerance and intolerance describe the users' sensitivity to changes in QoS parameter values. We next describe users' tolerance to latency and distortion.

- *Latency tolerance and intolerance:* This characteristic determines the stringency of the delay requirement. As we have discussed above, streaming multimedia applications are more latency tolerant than interactive multimedia applications. The degree of latency tolerance depends on users satisfaction, users expectation, or the urgency of the application (i.e., remote machine control in the manufacturing production line is latency intolerant).
- *Distortion tolerance and intolerance:* The tolerance to the fidelity of the application quality depends on factors such as users' satisfaction, users' expectation, and the application media types. For example, users are more tolerant to video distortion than to audio distortion. In this case, during congestion, the network has to maintain the quality of the audio output over the quality of the video output.

2.3.3 Adaptive and Nonadaptive Characteristics

Adaptive and nonadaptive aspects mostly describe the mechanisms invoked by the applications to adapt to QoS degradation. The common adaptive techniques are rate adaptation and delay adaptation:

- *Rate adaptive application* can adjust the data rate injected into the network. During network congestion, the applications reduce the data rate by dropping some packets, increasing the codec data compression, or changing the multimedia properties (i.e., reducing the resolution or color depth of the video). These schemes will most likely cause degradation of the perceived quality but will keep it within acceptable levels.
- *Delay-tolerant adaptive applications* tolerate a certain level of delay jitter by deploying the de-jittered buffer or adaptive playback technique (Figure 2.5).

Adaptation is trigged by some form of implicit or explicit feedback from the network or end user. Again, like tolerance, adaptive applications provide some room for networks to readjust the QoS services.

2.3.4 Application Criticality

Mission-critical and non-mission-critical aspects reflect the importance of application usage, which determines the strictness of the QoS requirements. Failing the mission may result in disastrous consequences. For example:

- *Remote surgery:* The surgeon performs an operation through remote surgical equipment. Life and death of the patient may depend on the promptness and accuracy of the surgical equipment control.
- *Telemedicine:* The accuracy of medical images (i.e., magnetic resonance image, x-ray image, ultrasound image) is extremely important. Distorted images may lead to wrong diagnosis.

2.3.5 Representation of Application Requirements

There are two common ways to express the applications' requirements: quantitative expression and qualitative expression.

2.3.5.1 Quantitative Expression

Application requirements are expressed in terms of QoS parameters with quantifiable values that can be determined from the application's technical specifications (i.e., video codec—MPEG 1, MPEG2, MPEG4, HDTV) or from experimentation. Some applications may obtain these values using runtime measurements.

- *Throughput:* It is mostly expressed as the average data rate:
  - Uncompressed HDTV: 1.5 Gbps
  - MPEG4: 5 kbps – 4 Mbps
  - ITU-T G.711: 64 kbps
  - ETSI GSM 06.10: 16 kbps

  For applications that generate VBR traffic, the average data rate does not properly capture the traffic characteristics. Burstiness parameters (average rate, peak rate, maximum burst size) may also be included.

- *Delay and delay jitter:* The value is provided in the form of a bound (i.e., the delay should be less than a certain value). In mission-critical or delay intolerant applications, the network has to follow the delay requirement strictly. In delay-tolerant applications, the bound value is the average value. The network can provide the service in a more relaxed manner (i.e., some of the packets can miss their deadline).
- *Loss:* It is mostly expressed as a statistical value (i.e., the percentage of lost packets should be less than a certain value).

2.3.5.2 Qualitative Expression

In contrast to the quantitative expression, QoS requirements are expressed qualitatively—for example, "…. get the bandwidth as much as you can give…" or " … expect low latency…" or " …. provide lower latency than certain applications…." There are a number of reasons why some applications (e.g., web browsing, remote login) express their QoS requirements qualitatively: some applications only need better service than others, or some applications are incapable of quantifying their QoS requirements, or it is too costly to quantify the QoS requirements. Since these applications do not specify quantitative QoS requirements, it is intractable for networks to provide quantitative services.

2.3.6 Examples of Application Requirements

2.3.6.1 Interactive Voice

Requirements for voice applications such as voice over IP (VoIP) can be exemplified as follows:

- *Bandwidth:* VoIP requires relatively low bandwidth (i.e., 64 kbps in ITU G.711) and the traffic pattern is relatively constant. Voice traffic is packetized in small packets of around 44 to 200 bytes resulting in short transmission times. Since voice conversations contain up to 60% silence, voice encoding algorithms yield lower bandwidth requirements with acceptable quality. Table 2.1 shows the speech coding standard. Mean Opinion Score (MOS) in Table 2.1 reflects the sound quality in a quantitative way. MOS is an empirical value determined by averaging the opinion score (i.e., the range from 1=worst to 5=best) of a sample group of listeners who listen and evaluate the voice quality.

Table 2.1 Speech Coding Standards

Codec	Bandwidth (kbps)	Sound Quality (MOS)	Codec Complexity
ITU-T G.711	64	> 4	—
ITU-T G.722	48-64	3.8	low
ITU-T G.723.1	6.4/5.3	3.9	high
ITU-T G.726	32	3.8	low
ITU-T G.728	16	3.6	low

Table 2.1 Speech Coding Standards (Continued)

Codec	Bandwidth (kbps)	Sound Quality (MOS)	Codec Complexity
ITU-T G.729	8	3.9	medium
GSM 06.10	13	3.5	low
GSM 06.20	5.6	3.5	high
GSM 06.60	12.2	> 4	high
GSM 06.70	4.8 – 12.2	> 4	high

- *Delay:* Since VoIP involves human interaction (conversation), human perception is sensitive to both the sound quality and the conversation gap and response. Table 2.2 displays delay guidelines for VoIP.

Table 2.2 Delay Guidelines for VoIP

One-Way Delay	Effect on Perceived Quality
< 100 – 150 ms	Excellent quality (undetectable delays)
150 – 250 ms	Acceptable quality (slight delays)
250 – 300 ms	Unacceptable quality

- *Delay jitter:* A large delay jitter causes delayed packets which impact the quality of the voice (conversation gap). Table 2.3 shows the delay jitter guidelines for VoIP.

Table 2.3 Delay Jitter Guidelines for VoIP

Delay Jitter	Effect on Perceived Quality
< 40 ms	Excellent quality—Undetectable jitter
40—75 ms	Acceptable quality
Over 75 ms	Unacceptable quality

- Not only the packet loss has impact on the voice quality, but also the pattern of the packet loss has impact. For example, the loss of one packet may not be noticeable due to the sophisticated codecs that can conceal the loss. However, two or more consecutive lost packets can cause voice quality deterioration.

2.3.6.2 Video Applications

Video applications can be classified into two groups: interactive video (i.e., video conferencing, long-distance learning, remote surgery) and streaming video (i.e., RealVideo, Microsoft ASF, QuickTime, Video on Demand, HDTV). As shown in Table 2.4, video applications' bandwidth requirements are relatively high depending on the video codec.

Table 2.4 Video Codec Bandwidth Requirements

Video Codec	Bandwidth
Uncompressed HDTV	1.5 Gbps
HDTV	360 Mbps
Standard definition TV (SDTV)	270 Mbps
Compressed MPEG2 4:4:4	25 – 60 Mbps
Broadcast quality HDTV (MPEG 2)	19.4 Mbps
MPEG 2 SDTV	6 Mbps
MPEG 1	1.5 Mbps
MPEG 4	5 kbps – 4 Mbps
H.323 (H.263)	28 kbps – 1 Mbps

- *Interactive video applications:* Interactive video which involves human interaction requires low end-to-end delay. Low delay jitter is also required especially in mission-critical applications such as remote surgery. Interactive video users are more tolerant to video distortion than streaming video users.

- *Streaming video applications:* Streaming video is considered to be one-way communication (from video server to users). The only interaction between the server and the user occurs during start, stop, forward, or reverse commands. Therefore, the delay is not a main concern. However, this application is delay jitter intolerant. Large delay jitter causes packets to miss the playback point, causing video distortion.

2.4 QoS Services

Networks receive from the applications (implicitly or explicitly) their QoS requirements through quantitative or qualitative expression. Networks need to respond to these requests by supplying QoS services using a number of QoS mechanisms. In this section, we describe various types of QoS services and in the next chapter we will introduce the mechanisms required to provide such services. In the rest of the book we will analyze each wireless technology and discuss which kind of services it can provide and which kind of applications it can support.

We categorize the QoS services as follows:

1. What kind of service is provided to applications: *quantitative services, qualitative services, best effort services*
2. To which entities (individual or group [class] of applications) the network provides service: *per-flow QoS services, per-class QoS services.*

Networks may use a combination of QoS services (i.e., per-flow and quantitative, per-class and quantitative). Some networks may include multiple types of QoS services in order to support a wide range of applications. Example of QoS services include the following:

- *Guaranteed Integrated Services (IntServ),* which provide per-flow and quantitative QoS service
- *Controlled Load Integrated Services,* which provide per-flow and qualitative QoS service
- *Differentiated Services (DiffServ),* which provide per-class and qualitative QoS service

We next describe in more detail each one of these QoS services.

2.4.1 Quantitative (Guaranteed) Services

Quantitative (guaranteed) services, or services that deliver hard QoS, guarantee the provision of the application quantitative requirements. This service delivers the highest quality of service. The guaranteed services guarantee the network performance (i.e., bandwidth, delay, delay jitter) in deterministic or statistical terms. For example, networks guarantee the minimum bandwidth provided to an application or guarantee the delay bound of packet delivery within a certain value. The services are suitable for applications that require quantitative performance guarantee such as mission-critical and interactive applications. As described in the next chapter, a number of QoS mechanisms are required to enable these services.

2.4.2 Qualitative (Differentiated) Services

Qualitative (differentiated) services, or services that deliver soft QoS, provide relative services. In other words, qualitative services may provide lower delay to one class of applications than to another class of applications. One prominent differentiated services example is the priority service. An application that belongs to a higher priority class will receive service before applications that belong to a lower priority class.

2.4.3 Best Effort Services

Best effort services provide network services without any performance guarantees. All traffic is treated equally. Best effort services are suitable for data traffic (i.e., FTP, email, web pages) that does not require minimum bandwidth or timed delivery.

2.4.4 Per-Flow QoS Services

Per-flow QoS services provide service assurance to individual flows (applications) quantitatively or qualitatively. For example, interactive video has stricter delay requirements than streaming video. Therefore, the network has to provide different services to each application in order to meet its individual needs. Per-flow classification (QoS mechanism that differentiates the flows) is essential to the implementation of per-flow QoS services.

2.4.5 Per-Class QoS Services

Applications are categorized into different groups (classes) based on different criteria (i.e., QoS requirement, organization, application types, protocol families). Per-class QoS services provide service assurance to individual classes quantitatively or qualitatively. Applications in the same class will experience the same QoS. The per-class classification (QoS mechanism that identifies and differentiates the collective entities) is essential to the implementation of per-class QoS services.

2.5 Realization of QoS Services

There are a large number of approaches, mechanisms, and technologies deployed in the network in order to enable the QoS services introduced in the previous section. The bandwidth, which is the main network resource, needs to be distributed to all applications in a way that simultaneously satisfies all QoS requirements. To enable QoS services, there are two main approaches (philosophies) based on how they deal with bandwidth planning in order to enable QoS services:

- *Bandwidth over-provisioning:* When the current network bandwidth cannot provide QoS support, the network infrastructure is upgraded. Using a higher bandwidth infrastructure may result in less congestion and therefore lower delivery delays.
- *Bandwidth management:* This approach proposes to manage the bandwidth using QoS mechanisms. Examples of QoS mechanisms which will be described in the next chapter are classification, admission control, resource reservation, channel access, packet scheduling, and policing.

There is an endless argument between the implementation of these two approaches. In wired networks, where bandwidth is abundant, over-provisioning is the winning approach. However, in wireless networks where the bandwidth is limited (spectrum is unavailable or is too expensive to purchase) bandwidth management techniques have to be deployed.



CHAPTER 3

QoS Mechanisms

3.1 Introduction

In the previous chapter, we introduced the fundamental QoS concepts. In this chapter we introduce a number of key QoS mechanisms that enable QoS services. At the end of this chapter, we provide a general framework for analyzing the QoS support of each wireless technology presented in the rest of this book.

QoS mechanisms can be categorized into two groups based on how the application traffic is treated: 1) traffic handling mechanisms, and 2) bandwidth management mechanisms (see Figure 3.1).

Traffic handling mechanisms (sometimes called In-traffic mechanisms) are mechanisms that classify, handle, police, and monitor the traffic across the network. The main mechanisms are: 1) classification, 2) channel access, 3) packet scheduling, and 4) traffic policing.

Bandwidth management mechanisms (sometimes called Out-of-traffic mechanisms) are mechanisms that mange the network resources (e.g., bandwidth) by coordinating and configuring network devices' (i.e., hosts, base stations, access points) traffic handling mechanisms. The main mechanisms are: 1) resource reservation signaling and 2) admission control.

3.2 Classification

The lowest service level that a network can provide is best effort service, which does not provide QoS support. In best effort service, all traffic is handled equally regardless of the application or host that generated the traffic. However, some applications need QoS support, requiring better than best effort service such as differentiated or guaranteed service. For a network to provide selective services to certain applications, first of all, the network requires a classification mechanism that can differentiate between the different applications. The classification mechanism identifies and separates different traffic into flows or groups of flows (aggregated flows or classes). Therefore, each flow or each aggregated flow can be handled selectively.

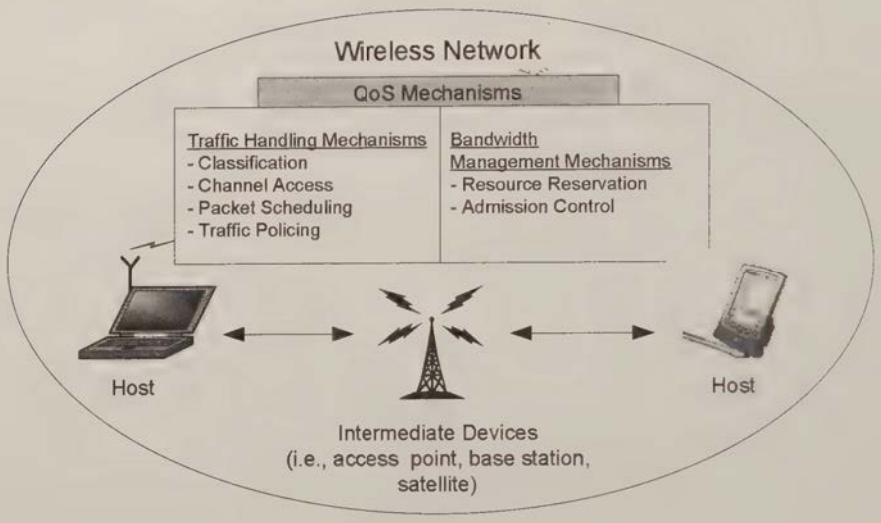


Figure 3.1 QoS Mechanisms in a Wireless Network

The classification mechanism can be implemented in different network devices (i.e., end hosts, intermediate devices such as switches, routers, access points). Figure 3.2 shows a simplified diagram of a classification module that resides on an end host and on an intermediated device.

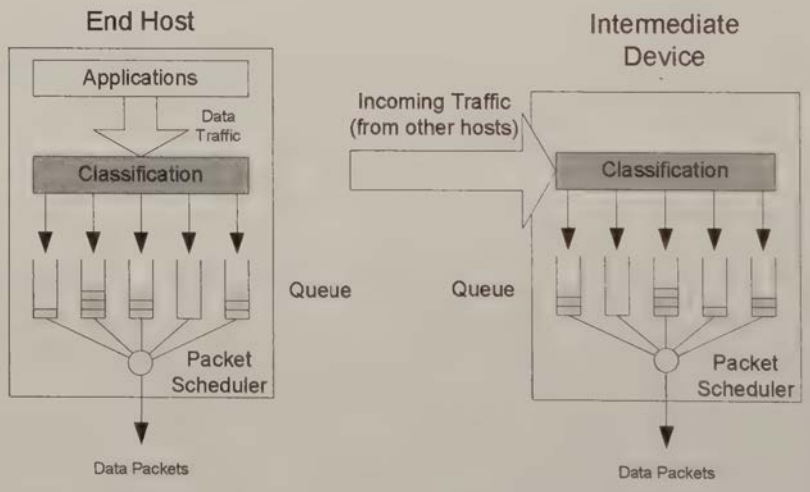


Figure 3.2 Classification

Application traffic (at the end host) or incoming traffic from other hosts (at the intermediate device) is identified by the classification mechanism and is forwarded to the appropriate queue awaiting service from other mechanisms such as the packet scheduler. The granularity level of the classification mechanism can be per-user, per-flow, or per-class depending on the type of QoS services provided. For example, per-flow QoS service requires per-flow classification while per-class QoS service requires per-class classification.

To identify and classify the traffic, the traffic classification mechanism requires some form of tagging or marking of packets. There are a number of traffic classification approaches. Some of approaches are suitable for end hosts and some for intermediate hosts. Figure 3.3 shows an example of some traffic classification approaches which are implemented in the different Open System Interconnection (OSI) layers.

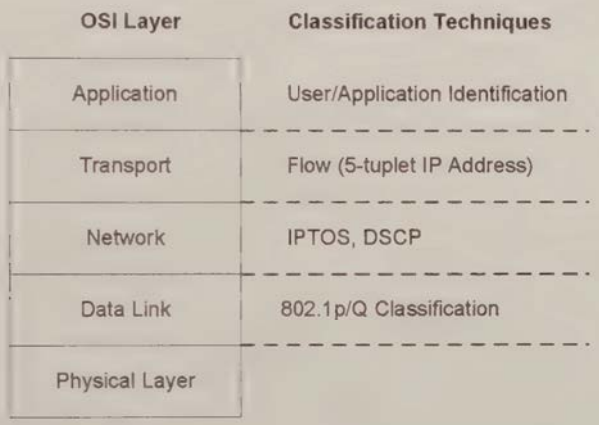


Figure 3.3 Examples of Existing Classification on Each OSI Layer

3.2.1 Data Link Layer Classification

Data link layer, or Layer 2, classifies the traffic based on the tag or field available in Layer 2 header.

An example of Layer 2 classification is IEEE (Institute of Electrical and Electronics Engineers) 802 user priority. The IEEE 802 header includes a 3-bit priority field that enables eight priority classes. It aims to support service differentiation on a Layer 2 network such as a LAN. The end host or intermediate host associates application traffic with a class (based on the Policy, or the service that the application expects to receive) and tags the packets' priority field in the IEEE 802 header. A classification mechanism identifies packets by examining the priority field of the IEEE 802 header and forwards the packets to the appropriated queues. IEEE recommends mapping the priority value and the corresponding service as shown in Table 3.1.

Table 3.1 Example of Mapping between Priority and Services

Priority	Service
0	Default, assumed to be best effort service
1	Less than best effort service
2	Reserved
3	Reserved
4	Delay sensitive, no bound
5	Delay sensitive, 100ms bound
6	Delay sensitive, 10ms bound
7	Network control

3.2.2 Network Layer Classification

Network layer, or Layer 3 classification, classifies packets using Layer 3 header. Layer 3 classification enables service differentiation in Layer 3 network.

An example of Layer 3 classification is IPTOS (Internet protocol type of service), DSCP (Internet protocol differential service code point). IPv4 and IPv6 standard defined a prioritization field in the IP header which can be used for Layer 3 classification. RFC 1349 defined a TOS field in IPv4 header. The type of service field consists of a 3-bit precedence subfield, a 4-bit TOS subfield, and the final bit which is unused and is set to be 0. The 4-bit TOS subfield enables 16 classes of service. In IPv6 header there is an 8-bit class of service field (see Figure 3.4). Later the Internet Engineering Task Force (IETF) differentiated services working group redefined IPv4 IPTOS to be DSCP, which is shown in Figure 3.4. DSCP has a 6-bit field enabling 64 classes of service.

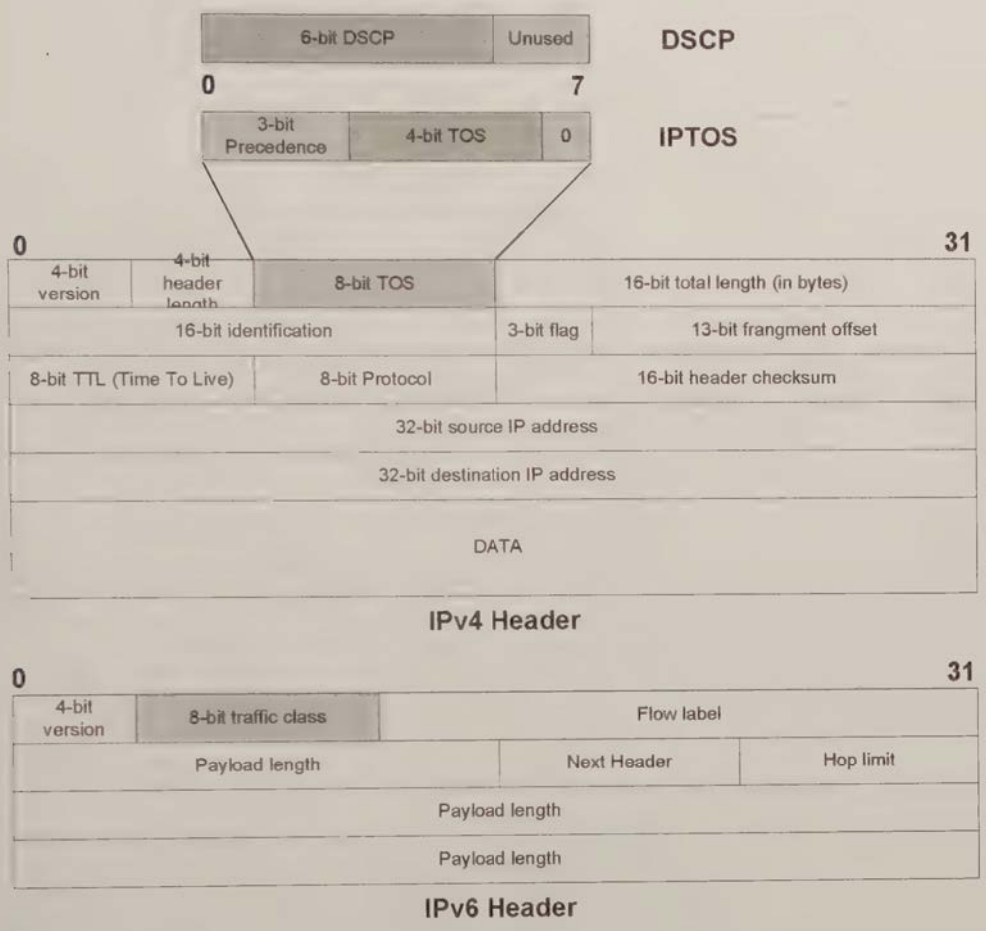


Figure 3.4 Structure of IPTOS and DSCP in IPv4 and IPv6

3.2.3 Transport Layer Classification (5-tuplet IP Header)

A 5-tuplet IP header (source IP, destination IP, source port, destination port, and protocol IP) can be used for transport layer classification. A 5-tuplet IP header can uniquely identify the individual application or flow. This classification provides the finest granularity and supports per-flow QoS service. However, the 5-tuplet IP header classification has some limitations:

- It is suitable for edge networks, but it is not suitable for core networks that carry very large amounts of traffic. Maintaining queues for each individual flow can be an overwhelming task.

- If the traffic passes through a firewall that uses NAT (network address translation), the real IP address (i.e., the IP address of the traffic source) is hidden from networks outside the firewall. Therefore, the 5-tuplet IP header exposed to a network outside the firewall cannot uniquely identify the application.

3.2.4 Application or User Classification

The application or user can be uniquely identified by using user/application identification (ID). The ID assignment may be static (i.e., the policy or the contract) or dynamic (i.e., connection signaling). For the connection signaling, there is a central station or entity in the network that is responsible for making the decision whether to allow a new session to join the network. First, the application or user sends the connection request to the central station. Then, if the new connection is admitted, it will be assigned a unique ID number. Packets from the application will be associated with an ID number.

3.3 Channel Access Mechanism

In wireless networks, all hosts communicate through a shared wireless medium. When multiple hosts try to transmit packets on the shared communication channel, collisions can occur. Therefore, wireless networks need a *channel access mechanism* which controls the access to the shared channel. There are two types of channel access mechanisms: 1) collision-based channel access and 2) collision-free channel access. Each type of channel access mechanism can provide different QoS services.

3.3.1 Collision-Based Channel Access

Collision-based channel access is a distributed channel access method that provides mechanisms to avoid collisions and to resolve collisions in case they occur. A classic collision-based channel access mechanism developed for wired LANs and implemented in Ethernet is CSMA/CD (Carrier Sense Multiple Access with Collision Detection). In collision-based channel access schemes, collisions can occur leading to the need for retransmissions. The collision probability depends on the number of active (with packets for transmission) users in the network. High traffic load increases the number of collisions and retransmissions, increasing the delay. Since we deal with stochastic traffic, the number of collisions and re-transmissions is random as well, leading to an unbound delay. Therefore, collision-based channel access schemes can provide best effort service. All hosts in the network receive equal bandwidth and experience the same unbounded delay. The service level can be improved by:

- Over-provisioning, whereby all traffic will receive ample of bandwidth and experience low delay.

- Adding a priority scheme in the collision-based channel access—that is, using different sized backoff windows for different priority classes. This will enable the provision of differentiated services. An example of such a solution is described in the proposed IEEE 802.11e (Chapter 4).

Existing solutions in wireless networks such as IEEE 802.11 DCF, HomeRF use collision-based channel access protocols similar to Ethernet CSMA/CD, denoted as CSMA/CA where CA stands for Collision Avoidance.

3.3.2 Collision-Free Channel Access

In a collision-free channel access mechanism the channel is arbitrated such that no collisions can occur. Only one host is allowed to transmit packets to the channel at any given time. Collision, therefore, will not occur. Examples of collision-free channel access techniques are polling and TDMA (Time Division Multiple Access).

3.3.2.1 Polling

A host in the network, or a specialized network device such as an Access Point or Base Station, is designated as the poller, which controls all access to the wireless channel by the other hosts denoted as pollees. Pollees are not allowed to transmit packets unless they receive a polling packet from the poller. Thus, there is no collision. Some pollees may receive the poll more often than others. The polling frequency (the number of polls in a period of time) reflects the bandwidth allocation. A poller can dynamically allocate bandwidth to pollees by adjusting the polling frequency dynamically.

3.3.2.2 TDMA (Time Division Multiple Access)

A TDMA scheme divides the channel access opportunity into frames and each frame is divided into time slots. A host is allowed to transmit packets in a predefined time slot, as shown in Figure 3.5.

The number of time slots assigned to a host per frame reflects the bandwidth allocated for the host. This technique requires a master host that is designated to manage the time slot assignment for all the hosts in the network. This Master host determines the number of time slots that each host will be allowed to transmit and notifies the hosts using some signaling mechanism. There are a number of time slot assignment philosophies:

- *Static time slot assignment:* Each host receives a fixed time slot assignment which can be provided during the connection setup.
- *Dynamic time slot assignment:* The time slot assignment changes dynamically during the lifetime of the session as a function of the traffic load, application QoS requirements, and channel conditions. This slot assignment policy is more flexible and leads to better channel utilization. However, it leads to signaling overhead required to communicate the slot assignment changes to the different hosts.

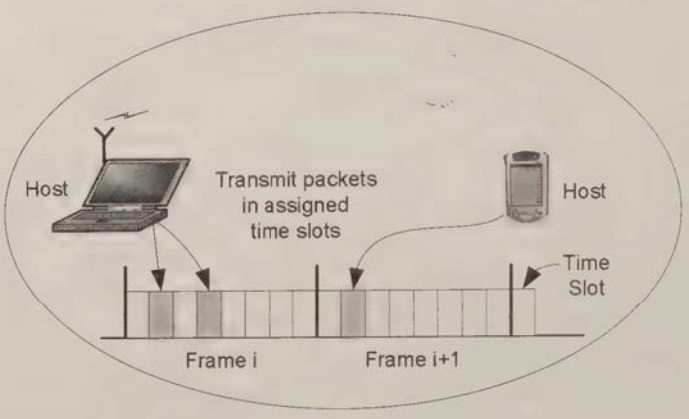


Figure 3.5 Time Division Multiple Access (TDMA) Scheme

Collision-free channel access schemes provide tight channel access control that can provide a tight delay bound. Therefore, these schemes are good candidates for QoS provision to applications with strict QoS requirements.

3.4 Packet Scheduling Mechanisms

Packet scheduling is the mechanism that selects a packet for transmission from the packets waiting in the transmission queue. It decides which packet from which queue and station are scheduled for transmission in a certain period of time. Packet scheduling controls bandwidth allocation to stations, classes, and applications.

As shown in Figure 3.6, there are two levels of packet scheduling mechanisms:

1. *Intrastation packet scheduling:* The packet scheduling mechanism that retrieves a packet from a queue within the same host.
2. *Interstation packet scheduling:* The packet scheduling mechanism that retrieves a packet from a queue from different hosts.

Packet scheduling can be implemented using hierarchical or flat approaches.

- *Hierarchical packet scheduling:* Bandwidth is allocated to stations—that is, each station is allowed to transmit at a certain period of time. The amount of bandwidth assigned to each station is controlled by interstation policy and module. When a station receives the opportunity to transmit, the intrastation packet scheduling module will decide which packets to transmit. This approach is scalable because interstation packet scheduling maintains the state by station (not by connection or application). Overall bandwidth is allocated based on stations (in fact, they can be groups, departments, or companies). Then, stations will have the authority to manage or allocate their own bandwidth portion to applications or classes within the host.

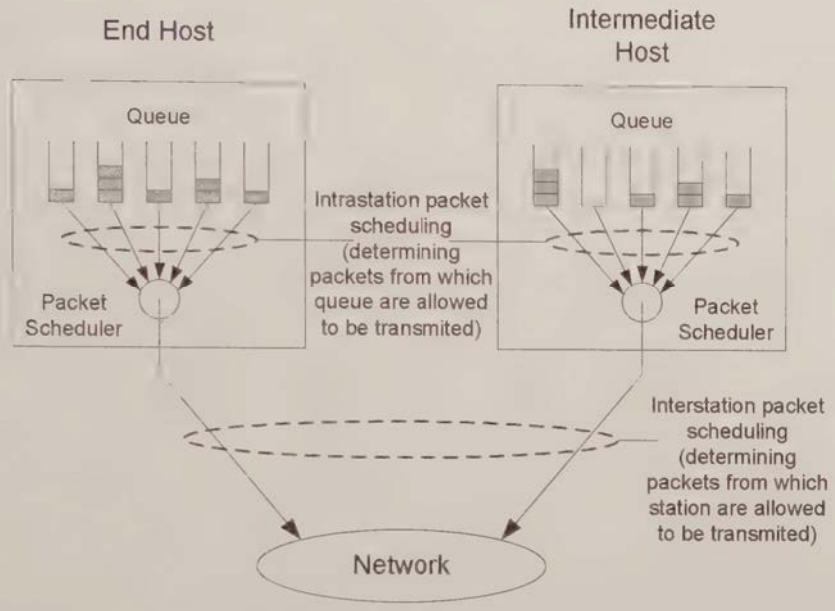


Figure 3.6 Packet Scheduling

- *Flat packet scheduling:* Packet scheduling is based on all queues of all stations. Each queue receives individual service from the network.

Packet scheduling mechanism deals with how to retrieve packets from queues, which is quite similar to a queuing mechanism. Since in intrastation packet scheduling the status of each queue in a station is known, the intrastation packet scheduling mechanism is virtually identical to a queuing mechanism. Interstation packet scheduling mechanism is slightly different from a queuing mechanism because queues are distributed among hosts and there is no central knowledge of the status of each queue. Therefore, some interstation packet scheduling mechanisms require a signaling procedure to coordinate the scheduling among hosts.

Because of the similarities between packet scheduling and queuing mechanisms we introduce a number of queuing schemes (First In First Out [FIFO], Strict Priority, and Weight Fair Queue [WFQ]) and briefly discuss how they support QoS services.

3.4.1 First In First Out (FIFO)

First In First Out (FIFO) is the simplest queuing mechanism. All packets are inserted to the tail of a single queue. Packets are scheduled in order of their arrival. Figure 3.7 shows FIFO packet scheduling.

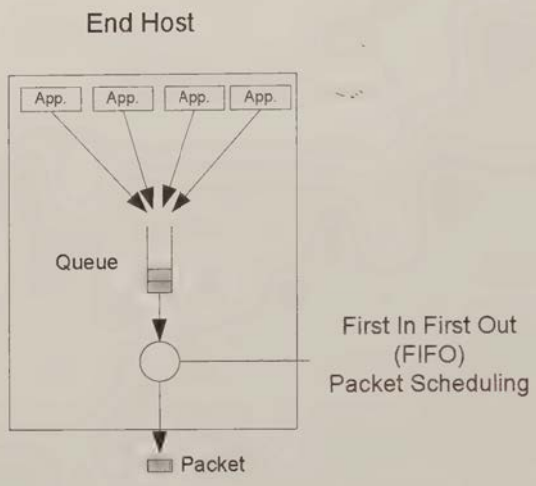


Figure 3.7 FIFO Packet Scheduling

FIFO provides best effort service—that is, it does not provide service differentiation in terms of bandwidth and delay. The high bandwidth flows will get a larger bandwidth portion than the low bandwidth flows. In general, all flows will experience the same average delay. If a flow increases its bandwidth aggressively, other flows will be affected by getting less bandwidth, causing increased average packet delay for all flows. It is possible to improve QoS support by adding 1) traffic policing to limit the rate of each flow and 2) admission control.

3.4.2 Strict Priority

Queues are assigned a priority order. Strict priority packet scheduling schedules packets based on the assigned priority order. Packets in higher priority queues always transmit before packets in lower priority queues. A lower priority queue has a chance to transmit packets only when there are no packets waiting in a higher priority queue. Figure 3.8 illustrates the strict priority packet scheduling mechanism.

Strict priority provides differentiated services (relative services) in both bandwidth and delay. The highest priority queue always receives bandwidth (up to the total bandwidth) and the lower priority queues receive the remaining bandwidth. Therefore, higher priority queues always experience lower delay than the lower priority queues. Aggressive bandwidth spending by the high priority queues can starve the low priority queues. Again, it is possible to improve the QoS support by adding 1) traffic policing to limit the rate of each flow and 2) admission control.

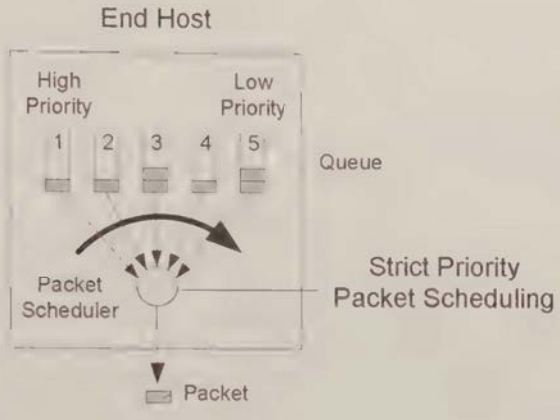


Figure 3.8 Strict Priority Packet Scheduling

3.4.3 Weight Fair Queue (WFQ)

Weight Fair Queue schedules packets based on the weight ratio of each queue. Weight, w_i, is assigned to each queue i according to the network policy. For example, there are three queues A, B, C with weights w_1, w_2, w_3, respectively. Queues A, B, and C receive the following ratios of available bandwidth: $w_1/(w_1+w_2+w_3)$, $w_2/(w_1+w_2+w_3)$, and $w_3/(w_1+w_2+w_3)$, respectively, as shown in Figure 3.9.

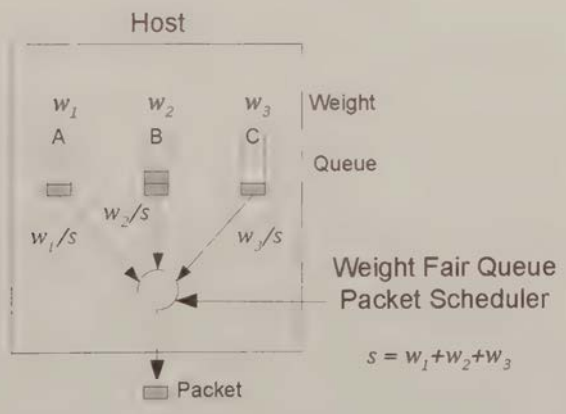


Figure 3.9 Weight Fair Queue Packet Scheduling

Bandwidth abuse from a specific queue will not affect other queues. WFQ can provide the required bandwidth and the delay performance is directly related to the allocated bandwidth. A queue with high bandwidth allocation (large weight) will experience lower delay. This may lead to some mismatch between the bandwidth and delay requirements. Some applications may require low bandwidth and low delay. In this case WFQ will allocate high bandwidth to these applications in order to guarantee the low delay bound. Some applications may require high bandwidth and high delay. WFQ still has to allocate high bandwidth in order for the applications to operate. Of course, applications will satisfy the delay but sometimes far beyond their needs. This mismatch can lead to low bandwidth utilization. However, in real life, WFQ mostly schedules packets that belong to aggregated flows, groups, and classes (instead of individual flows) where the goal is to provide link sharing among groups. In this case delay is of less concern.

The elementary queuing mechanisms introduced above will be the basis of a number of packet scheduling variations.

Before we move our discussion to the next QoS mechanisms, it is worth mentioning that in some implementations the channel access mechanism and packet scheduling mechanism are not mutually exclusive. There is some overlap between these two mechanisms and sometimes they are blended into one solution. When we discuss QoS support of each wireless technology in later chapters, in some cases, we will discuss both mechanisms together.

3.5 Traffic Policing Mechanism

Traffic policing is the mechanism that monitors the admitted sessions' traffic so that the sessions do not violate their QoS contract. The traffic policing mechanism makes sure that all traffic that passes through it will conform to agreed traffic parameters. When violation is found (e.g., more traffic is sent than was initially agreed upon in the QoS contract), a traffic policing mechanism is enforced by shaping the traffic. Because traffic policing shapes the traffic based on some known quantitative traffic parameters, multimedia (real-time) applications are naturally compatible to traffic policing. Most multimedia application traffic (voice, video) is generated by a standard codec which generally provides certain knowledge of the quantitative traffic parameters. Traffic policing can be applied to individual multimedia flows. Non-real-time traffic does not provide quantitative traffic parameters and usually demands bandwidth as much as possible. Therefore, traffic policing enforces non-real-time traffic (i.e., limits the bandwidth) based on the network policy. Such policing is usually enforced on aggregated non-real-time flows. Traffic policing can be implemented on end hosts or intermediate hosts. Examples of traffic policing mechanisms include the leaky bucket and the token bucket.

3.5.1 Leaky Bucket

The leaky bucket mechanism is usually used to smooth the burstiness of the traffic by limiting the traffic peak rate and the maximum burst size. This mechanism, as its name describes, uses the analogy of a leaky bucket to describe the traffic policing scheme. The bucket's parameters such as its size and the hole's size are analogous to the traffic policing parameters such as the

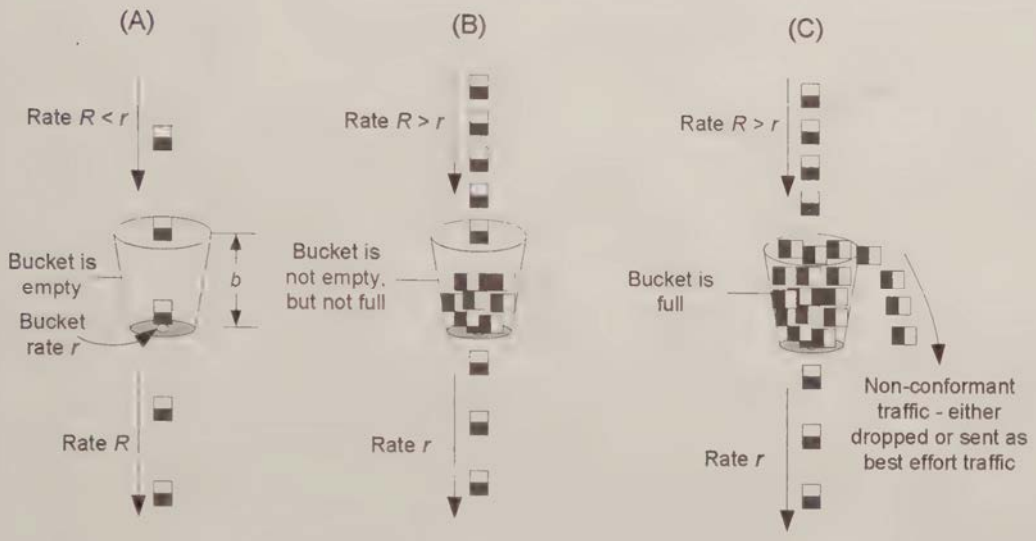


Figure 3.10 Leaky Bucket Mechanism

maximum burst size and maximum rate, respectively. The leaky bucket shapes the traffic with a maximum rate of up to the bucket rate. The bucket size determines the maximum burst size before the leaky bucket starts to drop packets.

The mechanism works in the following way. The arriving packets are inserted at the top of the bucket. At the bottom of the bucket, there is a hole through which traffic can leak out at a maximum rate of r bytes per second. The bucket size is b bytes (i.e., the bucket can hold at most b bytes). Let us follow the leaky bucket operation by observing the example shown in Figure 3.10. We assume first that the bucket is empty.

- Figure 3.10 (A): Incoming traffic with rate R which is less than the bucket rate r. The outgoing traffic rate is equal to R. In this case when we start with an empty bucket, the burstiness of the incoming traffic is the same as the burstiness of the outgoing traffic as long as $R < r$.
- Figure 3.10 (B): Incoming traffic with rate R which is greater than the bucket rate r. The outgoing traffic rate is equal to r (bucket rate).
- Figure 3.10 (C): Same as (B) but the bucket is full. Non-conformant traffic is either dropped or sent as best effort traffic.

3.5.2 Token Bucket

The token bucket mechanism is almost the same as the leaky bucket mechanism but it preserves the burstiness of the traffic. The token bucket of size b bytes is filled with tokens at rate r (bytes per second). When a packet arrives, it retrieves a token from the token bucket (given such a token is available) and the packet is sent to the outgoing traffic stream. As long as there are

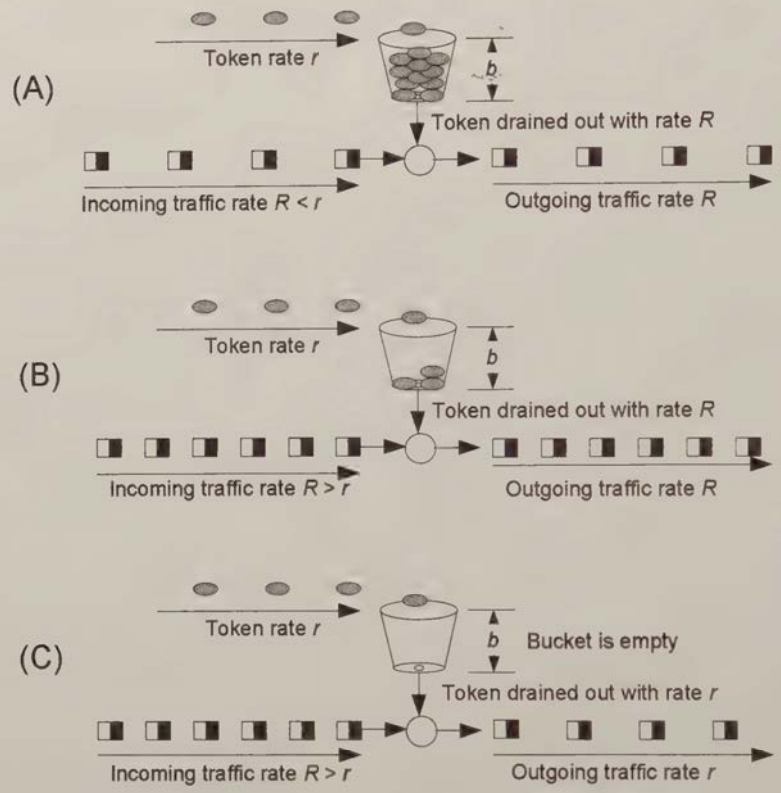


Figure 3.11 Token Bucket Mechanism

tokens in the token bucket, the outgoing traffic rate and pattern will be the same as the incoming traffic rate and pattern. If the token bucket is empty, incoming packets have to wait until there are tokens available in the bucket, and then they continue to send. Figure 3.11 shows an example of the token bucket mechanism.

- Figure 3.11 (A): The incoming traffic rate is less than the token arrival rate. In this case the outgoing traffic rate is equal to the incoming traffic rate.
- Figure 3.11 (B): The incoming traffic rate is greater than the token arrival rate. In case there are still tokens in the bucket, the outgoing traffic rate is equal to the incoming traffic rate.
- Figure 3.11 (C): If the incoming traffic rate is still greater than the token arrival rate (e.g., long traffic burst), eventually all the tokens will be exhausted. In this case the incoming traffic has to wait for the new tokens to arrive in order to be able to send out. Therefore, the outgoing traffic is limited at the token arrival rate.

The token bucket preserves the burstiness of the traffic up to the maximum burst size. The outgoing traffic will maintain a maximum average rate equal to the token rate, r. Therefore, the token bucket is used to control the average rate of the traffic.

In practical traffic policing, we use a combination of the token bucket and leaky bucket mechanisms connected in series (token bucket, then leaky bucket). The token bucket enforces the average data rate to be bound to token bucket rate while the leaky bucket (p) enforces the peak data rate to be bound to leaky bucket rate. Traffic policing, in cooperation with other QoS mechanisms, enables QoS support.

3.6 Resource Reservation Signaling Mechanisms

The traffic handling mechanisms (classification, channel access, packet scheduling, and traffic policing) we already described enable QoS services in each device. However, coordination between devices is essential to deliver end-to-end QoS services. Resource reservation signaling mechanisms inform the network entities on the QoS requirements of the multimedia applications using the network resources. The network devices will use this information to manage the network resources (i.e., bandwidth) in order to accommodate such requirements. The network devices control the network resources and provide QoS services by configuring the traffic handling mechanisms. Resource reservation can be applied to individual flows or aggregated flows. Resource reservation closely cooperates with the admission control mechanisms that will be described in a later section. Figure 3.12 shows a schematic diagram that describes the coordination between these mechanisms.

The resource reservation mechanisms include the following functions:

- Provision of resource reservation signaling that notifies all devices along the communication path on the multimedia applications' QoS requirements.
- Delivery of QoS requirements to the admission control mechanism that decides if there are available resources to meet the new request QoS requirements.
- Notification of the application regarding the admission result.

Resource Reservation Protocol (RSVP) is a well-known resource reservation signaling mechanism. RSVP operates on top of IP, in the transport layer, so it is compatible with the current TCP/IP based mechanisms (i.e., IPv4, IP routing protocol, and IP multicast mechanism) and can run across multiple networks. RSVP's main functionality is to exchange QoS requirement information among the source host, the destination host, and intermediate devices. Using this information, each network device will reserve the proper resources and configure its traffic handling mechanisms in order to provide the required QoS service. Once the reservation process is complete, the sender host is allowed to transmit data with an agreed traffic profile. If a device or network element on the communication path does not have enough resources to accommodate the traffic, the network element will notify the application that the network cannot support this QoS requirement. In order to achieve end-to-end resource reservation, all the network elements along

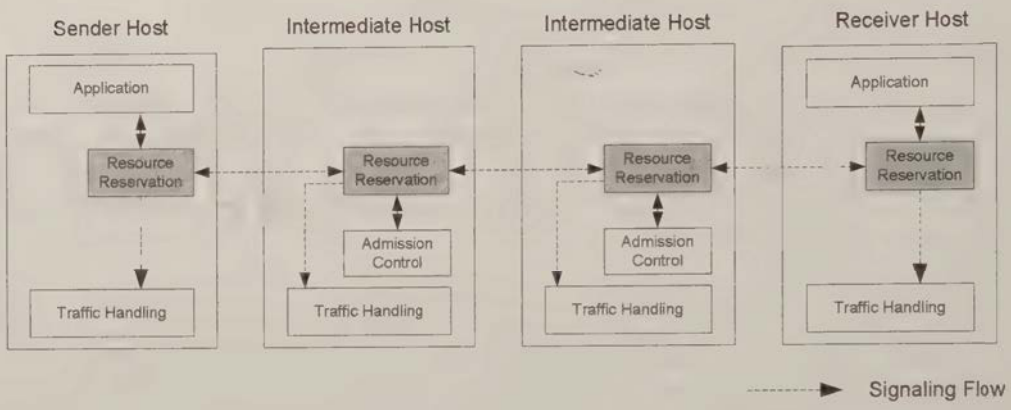


Figure 3.12 Resource Reservation Mechanism

the path (source host, destination host, and routers) need to be RSVP-enabled. Originally, RSVP was designed for supporting per-flow reservation. Currently it is extended to support per-aggregate reservation.

3.7 Admission Control

Admission control is the mechanism that makes the decision whether to allow a new session to join the network. This mechanism will ensure that existing sessions' QoS will not be degraded and the new session will be provided QoS support. If there are not enough network resources to accommodate the new sessions, the admission control mechanism may either reject the new session or admit the session while notifying the user that the network cannot provide the required QoS. Admission control and resource reservation signaling mechanisms closely cooperate with each other. Both are implemented in the same device. There are two admission control approaches:

- *Explicit admission control:* This approach is based on explicit resource reservation. Applications will send the request to join the network through the resource reservation signaling mechanism. The request that contains QoS parameters is forwarded to the admission control mechanism. The admission control mechanism decides to accept or reject the application based on the application's QoS requirements, available resources, performance criteria, and network policy.
- *Implicit admission control:* There is no explicit resource reservation signaling. The admission control mechanism relies on bandwidth over-provisioning and traffic control (i.e., traffic policing).

The location of the admission control mechanism depends on the network architecture. For example, in case we have a wide area network such as a high-speed backbone that consists of a number of interconnected routers, the admission control mechanism is implemented on each router. In shared media networks, such as wireless networks, there is a designated entity in the

network (e.g., station, access point, gateway, base station) that hosts the admission control agent. This agent is in charge of making admission control decisions for the entire wireless network. This concept is similar to the SBM (subnet bandwidth manager) which serves as the admission control agent in 802 networks.

In ad hoc wireless networks, the admission control functionality can be distributed among all hosts. In infrastructure wireless networks where all communication passes through the access point or base station, the admission control functionality can be implemented in the access point or base station.

3.8 QoS Architecture

This section shows how all the QoS mechanisms described in the previous subsections are working together to provide QoS support. Different applications that co-exist in the same network may require different combinations of QoS mechanisms such as:

- *Applications with quantitative QoS requirements:* These applications mostly require QoS guaranteed services. Therefore, explicit resource reservation and admission control are needed. They also require strict traffic control (traffic policing, packet scheduling, and channel access).
- *Applications with qualitative QoS requirements:* These applications require high QoS levels but do not provide quantitative QoS requirements. In this case we can use resource reservation and admission control. They also require traffic handling which delivers differentiated services.
- *Best effort:* There is no need for QoS guarantees. The network should reserve bandwidth for such services. The amount of reserved bandwidth for best effort traffic is determined by the network policy.

The QoS architecture which contains different QoS mechanisms is different for each network topology. We will focus on the QoS architecture for ad hoc and infrastructure wireless networks.

3.8.1 QoS Architecture for Infrastructure Wireless Networks

In infrastructure wireless networks, there are two types of stations: end stations (hosts) and a central station (i.e., access point, base station). The central station regulates all the communication in the network—that is, there is no peer-to-peer communication that occurs directly between the hosts. The traffic from a source host is sent to the central station and then the central station forwards the traffic to the destination host. All traffic handling (classification, traffic policing, packet scheduling, and channel access) and resource reservation mechanisms reside in all stations (end hosts and central station). In addition, the central station also includes an admission control mechanism. Figure 3.13 shows a QoS architecture for an infrastructure wireless network.

There are some variations in the signaling mechanisms that configure the traffic handling mechanisms in each station. We will point out these differences in each wireless technology chapter.

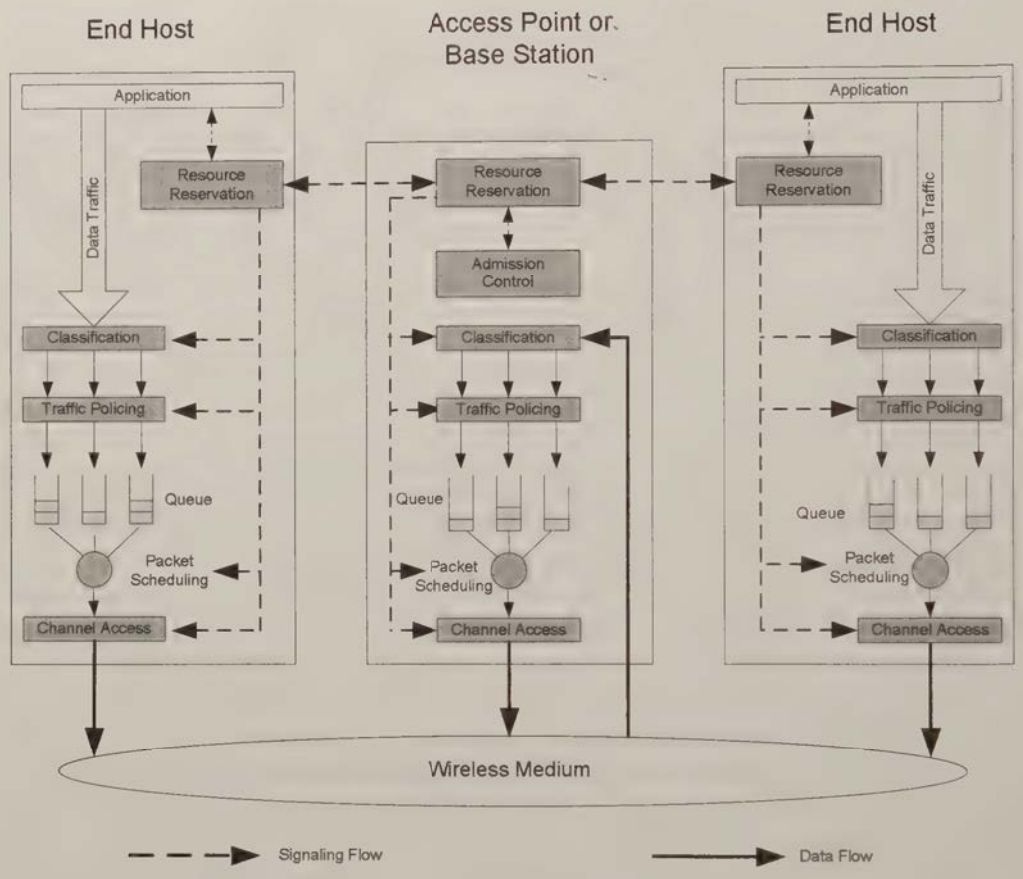


Figure 3.13 QoS Architecture of an Infrastructure Wireless Network

3.8.2 QoS Architecture For Ad Hoc Wireless Networks

All hosts establish peer-to-peer communication in the shared wireless media environment. All traffic handling and resource reservation mechanisms reside in all hosts. One of the hosts (either a dedicated or a regular end host) will be designated to serve as an admission control agent (i.e., designated SBM [DSBM]). Figure 3.14 shows a QoS architecture for an ad hoc wireless network.

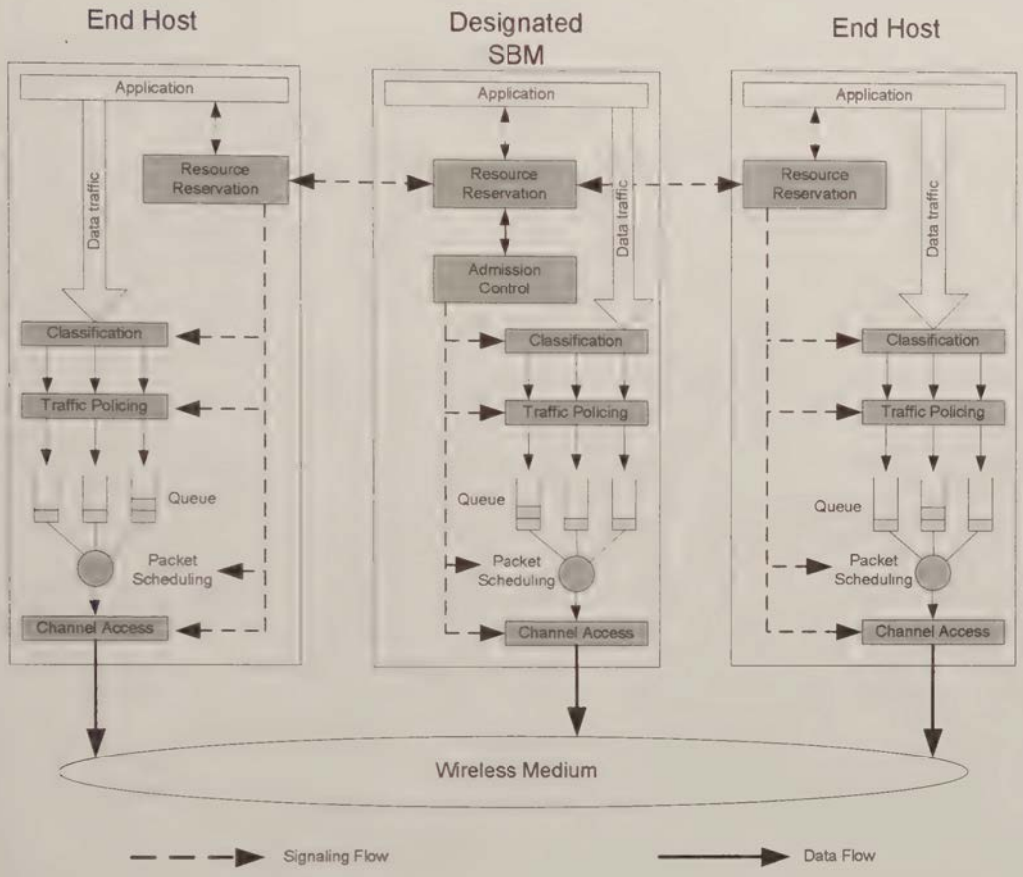


Figure 3.14 QoS Architecture for an Ad Hoc Wireless Network



PART 2

Wireless Local Area Networks





CHAPTER 4

IEEE 802.11

4.1 IEEE 802.11

As shown in Figure 4.1, the IEEE (Institute of Electrical and Electronics Engineers) standards committee has defined a family of IEEE 802 LAN (local area network) and MAN (metropolitan area network) standards in which IEEE 802.11 defines the wireless LAN (WLAN) standard with multiple physical layer options including speeds of up to 54 Mbps. The organization is expected to introduce enhancements to the standard that will support QoS and other radio options. Many companies (e.g., Cisco, 3Com, Linksys, D-Link) have introduced communication cards based on this standard.

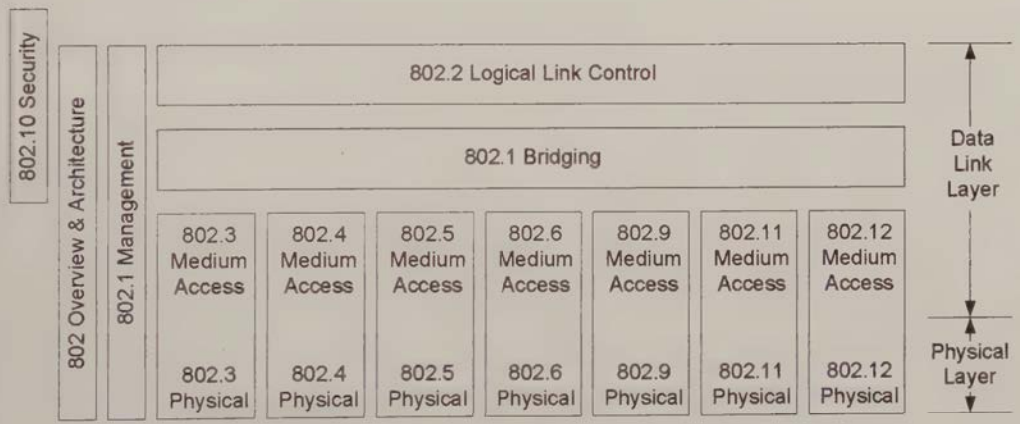


Figure 4.1 IEEE 802 Standards

Compatibility and interoperability among various IEEE 802.11 vendors are managed by the Wireless Ethernet Compatibility Alliance (WECA). Their certification is referred as Wi-Fi, which stands for Wireless Fidelity. Wi-Fi certification gives consumers and businesses the assurance that WLAN products bearing the Wi-Fi logo are compatible and work together even when manufactured by different vendors. Such PC products include PCMCIA (PC Memory Card International Association) cards for notebooks, PCI (peripheral component interconnect/interface) cards for desktops, and cards with USB (Universal Serial Bus) interface that can be used with either one.

The wireless media are fundamentally different from the wired media in the following aspects:

- The wireless media have no clear physical boundaries in which users can transmit and receive data.
- Users can sometimes be covered by multiple WLANs. This fact requires a resolution mechanism that will enable the users to conduct effective communication with minimal interruptions from other WLANs' participants.
- Users are susceptive to radio interferences coming from sources such as microwaves, causing the wireless media to be significantly less reliable then the wired media.
- Users are mobile—that is, they can change their locations, and roam within their coverage areas. This mobility causes a dynamic network topology, which leads to the need to have procedures for the users to associate/disassociate with the specific WLAN. Moreover, a user can be temporarily hidden from some other network members.
- The power limitation, the wireless media nature, and the topographic characteristics of the communication area impose a limited range for the effective geographical coverage. This fact may require the installation of several overlapping wireless LANs to cover a larger geographical area.
- Many, if not most, of the wireless devices are battery powered. Hence, power saving is essential to increase the time between battery charges. This fact may cause situations in which the user is temporarily without service.

The IEEE 802.11 family consists of a few standards. The first standard that was approved in 1997 details the Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications that support data rates of 1 Mbps and 2 Mbps over the 2.4 GHz Industrial, Scientific, and Medical (ISM) frequency band using either Frequency Hopping Spread Spectrum (FHSS) or Direct Sequence Spread Spectrum (DSSS) as their radio technologies as well as infrared. (Note: ISM frequency bands are nonlicensed frequency bands used for wireless network, cordless phone, and other devices. ISM frequency bands include 902–928 MHz, 2.4–2.5 GHz and 5.725–5.875 GHz.) The standard details operational specifications for wireless connectivity for fixed, ad hoc, and mobile stations within a local geographical area. This first standard was introduced after years of careful development and has become popular with vendors who implemented significant parts of this standard.

The IEEE 802.11 standard contains two additions introduced later: IEEE 802.11a enhances the speed to up to 54 Mbps using Orthogonal Frequency Division Multiplexing (OFDM) radio technology in the 5 GHz band, and IEEE 802.11b enhances the speed to 5.5 Mbps and 11 Mbps using Complementary Code Keying (CCK) modulation in the 2.4 GHz band. These standards allow for multirate support and backwards compatibility (i.e., IEEE 802.11b supports speeds of 1, 2, 5.5, and 11 Mbps). As long as users in the same WLAN use the same basic radio technology, even though with different data rates, these users will be able to communicate using the lowest speed. For example, a lower speed DSSS based radio (i.e., IEEE 802.11) will be able to communicate with a more advanced CCK based radio (i.e., IEEE 802.11b) sharing the lower speed (1 or 2 Mbps) of the two radios. The IEEE 802.11g committee is working on the definition of a standard in which higher speeds of more than 20 Mbps will be used within the 2.4 GHz band. This potential extension will be backwards compatible and interoperable with IEEE 802.11b. IEEE 802.11g provides speeds of 1, 2, 5.5, 6, 9, 11, 12, 18, 24, 36, 48, and 54 Mbps.

The IEEE 802.11 standards consider data traffic but lack proper consideration for multimedia applications' QoS needs. Currently, the IEEE 802.11 group is working on an extension, IEEE 802.11e, that will specify how signaling and support for QoS can be achieved.

Furthermore, the IEEE 802.11 group is also working on a number of standard supplements as summarized in Table 4.1.

Table 4.1 IEEE 802.11 Standard Supplements

Standard Supplements	Brief Description
IEEE 802.11a	Define a PHY in newly allocated UNII band. The standard has been completed and published as 8802-11: 1999 (E)/Amd 1: 2000 (ISO/IEC) (IEEE Std. 802.11a-1999 Edition).
IEEE 802.11b	Define a high rate PHY in the 2.4 GHz band. The standard has been completed and published as IEEE Std. 802.11b-1999.
IEEE 802.11c	Define MAC procedure for the bridge operation. The standard has been completed and is now part of IEEE 802.11c standard.
IEEE 802.11d	Define PHY requirements to extend the operation of 802.11 WLAN to new regulatory domains (countries). The standard has been completed and published as IEEE Std. 802.11d-2001.
IEEE 802.11e	Enhance current 802.11 MAC to expand support for applications with QoS requirement.

Table 4.1 IEEE 802.11 Standard Supplements (Continued)

Standard Supplements	Brief Description
IEEE 802.11f	Recommend practice for Inter Access Point Protocol (IAPP).
IEEE 802.11g	Standard for high-speed (20+ Mbps) PHY extensions to the 802.11b standard.
IEEE 802.11h	Enhance 802.11 MAC and 802.11a PHY with network management, control extensions for spectrum and transmit power management in 5GHz license exempt bands.
IEEE 802.11i	Enhance the 802.11 MAC to enhance security and authentication mechanisms.
IEEE 802.11j	Enhance the current 802.11 MAC and 802.11a PHY to additionally operate in newly available Japanese 4.9 GHz and 5 GHz bands.
IEEE 802.11k	Define Radio Resource Measurement enhancements to provide interfaces to higher layers for radio and network measurements.

IEEE 802.11a, b, c, and d have already been approved, whereas IEEE 802.11e, f, g, h, i, j, and k are under development.

4.1.1 Architecture

4.1.1.1 Network Topologies

The standard addresses two topologies: ad hoc topology, referred to as Independent Basic Service Set (IBSS), and infrastructure topology, referred to as Basic Service Set (BSS). A topology that combines several BSS cells is referred to as Extended Service Set (ESS).

4.1.1.1.1 Ad Hoc Network As shown in Figure 4.2, an ad hoc network or an IBSS consists of stations within mutual communication range of each other via the wireless medium. Such a network is created spontaneously, without preplanning, for ad hoc temporary situations with limited needs to access the Internet. The IBSS is the most basic type of an IEEE 802.11 WLAN and may contain only two stations. Figure 4.2 shows two IBSSs, each with two stations. If a station moves out of its IBSS, meaning out of range, it can no longer communicate with the other IBSS members.

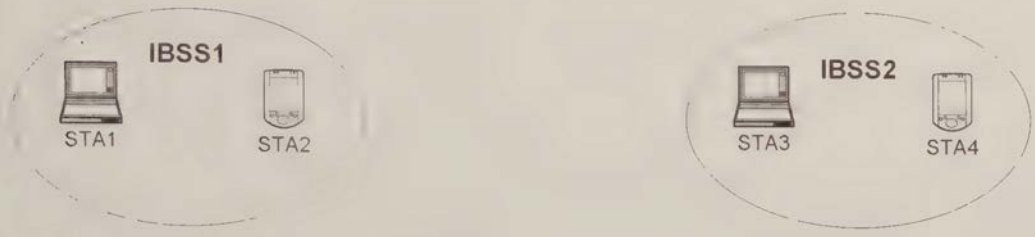


Figure 4.2 Ad Hoc Topology: Independent Basic Service Sets

4.1.1.1.2 Infrastructure Network The infrastructure network or BSS includes an access point (AP) in addition to the stations. This AP acts as the BSS arbitrator, meaning that the AP will handle all the BSS traffic. The BSS traffic can be either internal traffic (i.e., among the BSS participants) or external traffic (between the BSS participants and outside the BSS). The AP integrates the BSS within the distribution network. For example, all traffic between the BSS participants and the Internet will be delivered through the AP.

Figure 4.3 shows an ESS composed of two BSSs, each with two stations and an AP. Each BSS is interconnected to the distribution system (DS) (which may connect to the Internet) through the AP. The key concept is that the ESS network appears as a single network entity to the upper layers including the applications.

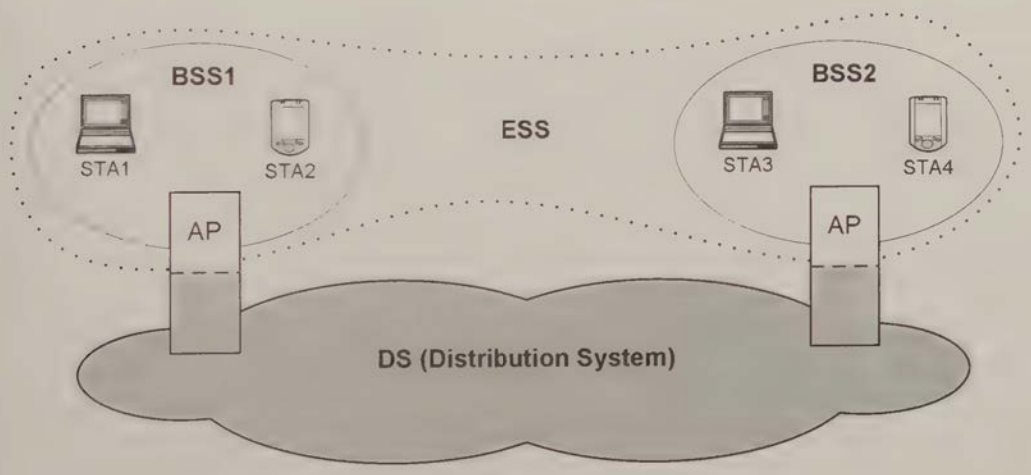


Figure 4.3 Infrastructure Topology: Extended Service Set

To allow effective communication for users who constantly move from one BSS geographical area to another BSS area, the BSSs should be physically overlapping. In the standard there is no limit to the distance between the BSSs and it is left up to the user or network installer to determine the BSS location. Moreover, other users may decide to install an IBSS in the same geographical area. In this case, multiple WLANs will be collocated and function effectively if the respective users coordinate the used frequencies.

4.1.1.1.3 Collocation Wireless networks can be located in overlapping geographical areas (see Figure 4.4). Radio signals may propagate from one network to another. A change in the transmitter position and geographical layout may have a profound effect on the propagation characteristics. Hence such networks do not have a clear boundary. The possibility of having multiple collocated operational WLANs not only allows multiple independent user groups to work simultaneously in the same place but also aggregates and increases the WLAN capacity.

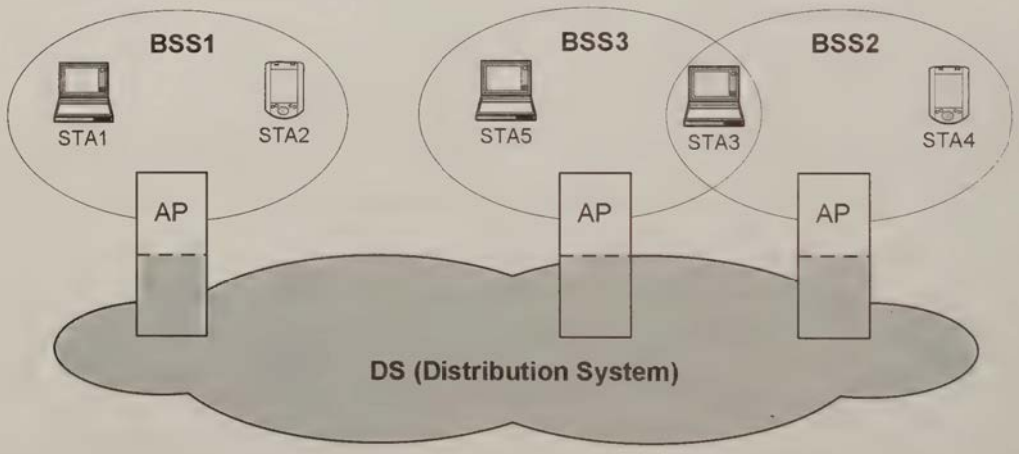
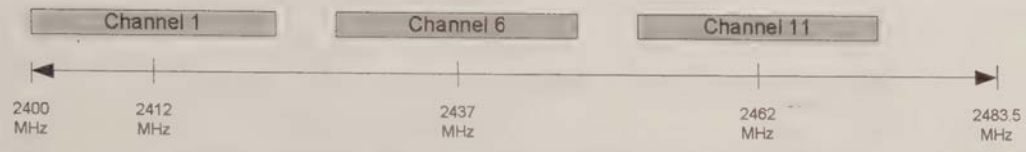


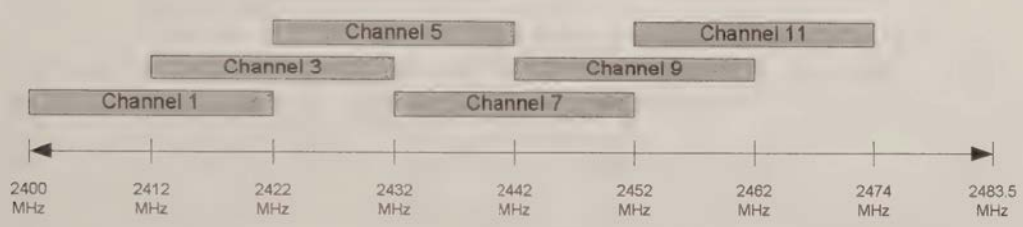
Figure 4.4 Collocated WLANs

Collocation is possible if network installers make an effort to coordinate the frequencies used by their respective WLANs. If they use DSSS based WLANs they can coordinate the center frequency. Several centers are possible according to the region in which the networks are installed. For example, IEEE 802.11b in North America details three non-overlapping channels and six overlapping channels (see Figure 4.5). This means that up to three networks can be located without mutual interruptions.

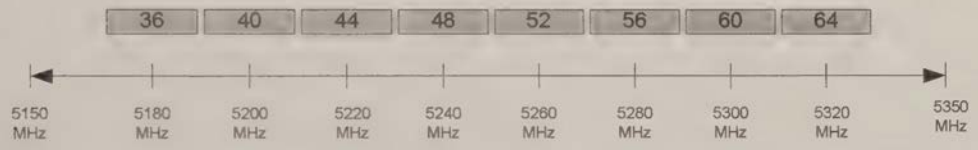
Collocation is also possible in FHSS-based WLANs. IEEE 802.11b defines 78 hopping sequences (each with 79 hops) grouped in three sets of 26 sequences each. This means that theoretically 26 FHSS WLAN systems can be collocated. However, practically fewer user groups can be collocated because of interference from other groups. This number is estimated to be 15. In IEEE 802.11a collocation is possible if different WLANs use different operating frequencies.



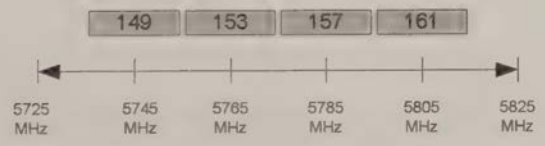
North American Channel Selection - non-overlapping (2.4 GHz ISM frequency band)



North American Channel Selection - overlapping (2.4 GHz ISM frequency band)



North American Channel Selection (Lower U-NII 5150 - 5250 MHz, Middle U-NII 5250 - 5350 MHz)



North American Channel Selection (Upper U-NII 5725 - 5825 MHz)

Figure 4.5 North American Channels

In the U.S., there are three 100 MHz unlicensed national information infrastructure (U-NII) bands: Lower U-NII (5.15 to 5.25 GHz), Middle U-NII (5.25 to 5.35 GHz), and Upper U-NII (5.725 to 5.825 GHz).

4.1.1.2 Protocol Stack

As shown in Figure 4.6, the IEEE 802.11-1997 base standard focuses on the following two parts of the protocol stack: the MAC part of the data link layer and the physical layer.

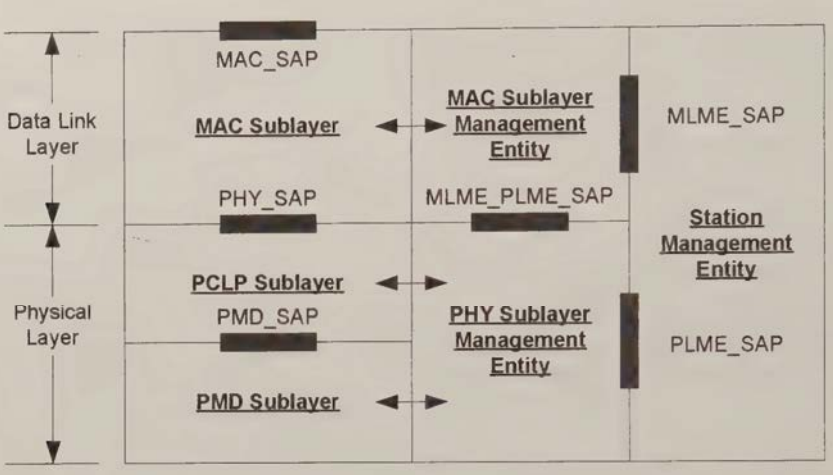


Figure 4.6 Portions of the OSI Protocol Stack Covered by IEEE 802.11 Base Standard (PLCP [physical layer convergence protocol]). Abbreviations: MLME—MAC Sublayer Management Entity; PLME—Physical Layer Management Entity; SAP—Service Access Point.

4.1.2 Physical Layer

IEEE 802.11 allows for various wireless technologies: infrared and radio DSSS, FHSS, and OFDM. One technology cannot work with the other. In other words, radio FHSS cannot communicate with a radio that employs DSSS. The standard, though, specifies backwards compatibility for radios that use the same transmission technology. In other words, a higher speed radio will be able to communicate with a lower speed radio via adjusting the speed of the higher speed radio to that of the lower speed radio. IEEE 802.11a allows for higher speeds in the 5 GHz band based on OFDM: 6, 9, 12, 18, 24, 36, 48, and 54 Mbps (support of 6, 12, and 24 Mbps data rates is mandatory). OFDM is also considered by IEEE 802.11g at the 2.4 GHz frequency range to achieve speeds of 1, 2, 5.5, 6, 9, 11, 12, 18, 24, 36, 48, and 54 Mbps.

The various radio flavors use slightly different frequencies in different countries and continents. This means that a U.S. radio will not work properly with a Japanese radio. Ongoing work is conducted to allow better interoperability and to minimize the need for equipment manufacturers to produce a wide variety of country-specific products and for users to travel with a bag full of country-specific WLAN cards. Such effort is reflected in the IEEE 802.11d that supplements the IEEE 802.11 with features that allow WLANs to operate within the rules of different countries. Another effort is IEEE 802.11f, whose goal is to achieve radio access point interoperability within a multivendor WLAN environment.

4.1.3 Media Access Control (MAC)

The MAC governs the stations' access to the shared wireless medium. The MAC which is located with the data link layer has a crucial role in providing QoS support to users, especially when executing multimedia applications.

The MAC architecture details two operational modes that coexist (see Figure 4.7): the Distributed Coordination Function (DCF) and the Point Coordination Function (PCF). The two modes use a cycle structure (see Figure 4.8) denoted collision-free period repetition interval, in which the first time period (called Contention-free Period or CFP) is governed by the PCF mode and the second time period (called Contention Period or CP) is governed by the DCF mode. DCF is a simplistic mode that allows contention and collision of traffic between stations. Therefore, DCF is suitable for applications that do not require QoS. PCF is a more complex protocol that uses the Point Coordinator (PC) which resides at the AP. The PC arbitrates the channel access using a polling based approach. PCF is geared to provide QoS support for multimedia applications. However, there are no known products that implement the PCF mode. QoS provisioning is going through an extensive review by the IEEE 802.11 organization under the extension IEEE 802.11e. This is discussed in Section 2 of this chapter.

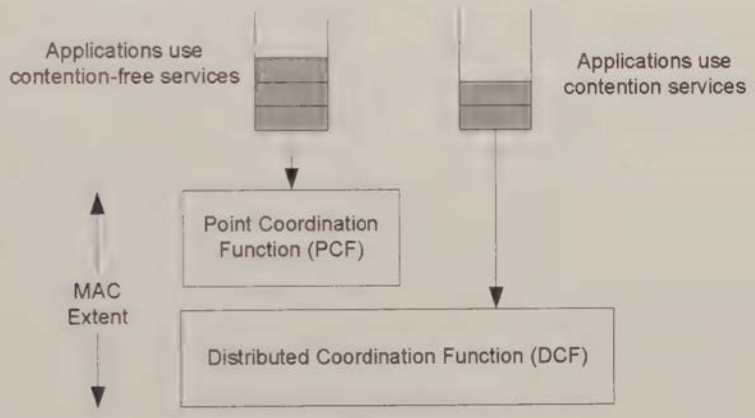


Figure 4.7 IEEE 802.11 MAC Architecture

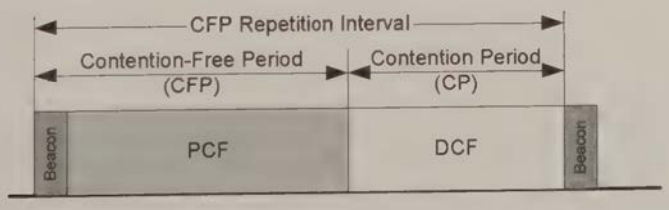


Figure 4.8 IEEE 802.11 Collision-Free Period Repetition Interval

In addition to defining the media access rules for each mode, the standard also defines procedures for traffic fragmentation and defragmentation, multirate support, authentication and privacy. Fragmentation, or slicing the original data frame into smaller transmitted frames, is carried in order to reduce the loss of traffic which can occur due to collisions with other users' traffic or due to radio interference. The smaller the frame, the less wireless media bandwidth is lost in the collision. However, smaller frames lead to more overhead consisting of frames headers that contain addresses and other pertinent information. Hence, there is a tradeoff which will determine how much fragmentation should be carried out. The standard does not quantify the level of fragmentation and leaves it to the implementers' discretion.

4.1.3.1 Distributed Coordination Function (DCF)

The fundamental channel access mechanism of the IEEE 802.11 MAC is DCF, also known as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). This is a random access mechanism that enables sharing of the wireless media between compatible radio transmitters. The standard mandates that the DCF is implemented in all stations for use within both the IBSS and BSS configurations.

In the DCF mode a station cannot transmit arbitrarily, but rather needs to go through a set of steps to determine whether it can transmit. The goal of these steps is to reduce the chance of collision with other stations' packets. Such collision causes the destruction of the transmitted packet along with other stations' packets. Before transmission, the user's transmitter senses the wireless medium to determine if another station is transmitting. If the wireless medium is determined to be free, the transmitter will sense the channel and wait before transmission for a certain time referred to as Interframe Space (IFS). Once the data are successfully received by the addressee, the addressee is required to respond with an immediate positive acknowledgment (ACK) such that the user is assured that the data have arrived. In case the ACK is not received, the user station will go through a process of retransmission until an ACK is received. When a retransmission is required, the transmitter needs to wait an additional random backoff time before transmitting in order to minimize collisions with the other stations that also participated in the collision.

IEEE 802.11 defines four IFSs with different duration (see Figure 4.9): Short Interframe Space (SIFS), PCF Interframe Space (PIFS), DCF Interframe Space (DIFS), and Extended Interframe Space (EIFS). The duration of these intervals is determined in the standard as a function of the radio technology used. The duration of these intervals determines the priority of the packets (i.e., in case the packet waits SIFS, which is shortest interframe interval, the packet has the highest priority).

The IFS intervals are used as follows:

- SIFS is used as an interframe space between a data frame and its ACK frame, between an RTS (request to send) frame and its CTS (clear to send) frame between fragments of the original data frame, and by a station responding to a control message in the PCF mode.

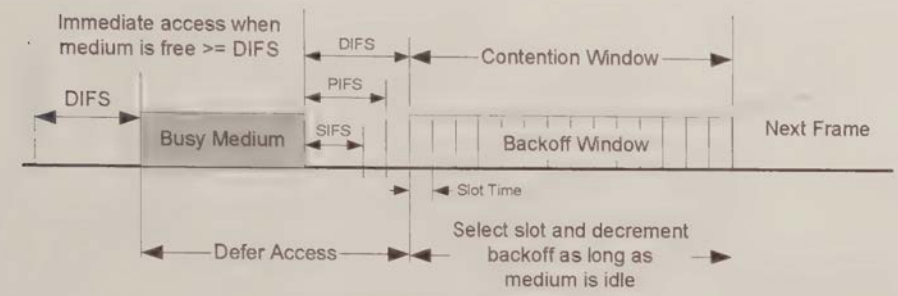


Figure 4.9 IEEE 802.11 Interframe Spacing (IFS)

- PIFS is used in the PCF mode to gain priority access to the medium.
- DIFS is used in the DCF mode to transmit data and management frames. If the wireless medium is free for the DIFS period the station is allowed to transmit.

Despite this elaborate mechanism collisions can occur if stations are far away from each other and cannot listen to each other's transmissions or if the stations start transmitting concurrently. The more stations there are in the BSS or IBSS the more collisions occur, and hence bandwidth utilization is deteriorated.

The standard provides a mechanism to further minimize collisions by reserving the channel for a certain period of time. During this period of time only one station is allowed to transmit while others defer. This reservation is obtained through the following process. The transmitting station (i.e., the station that would like to reserve the channel) transmits a special announcement to all stations, announcing its intent to transmit in the period mentioned in this announcement. The addressed station is expected to respond and acknowledge this announcement. The announcement is referred to as Request To Send (RTS) packet and the addressee's response is referred to as Clear To Send (CTS) packet. All stations who receive either the RTS or CTS will refrain from their transmission attempts until the end of the time period announced in either the RTS or CTS packets. The RTS/CTS mechanism cannot be used for broadcast and multicast traffic. The RTS/CTS burdens the network with additional traffic and overhead that reduces some of its benefit. The standard does not define when and in what situations the RTS and CTS exchange should be carried out. This decision is left to the implementer and even the user in cases where the wireless communication card provides access to this decision.

Based on the CSMA/CA protocol, each station maintains a prediction of time in which future traffic will occur. This prediction mechanism is referred in the standard as the Network Allocation Vector (NAV). This duration is announced either in the RTS/CTS packets or in the header of the packets sent during the DCF mode. The stations use the NAV to defer from accessing the wireless medium.

4.1.3.2 Point Coordination Function (PCF)

The IEEE 802.11 MAC standard defined an optional access method called Point Coordination Function (PCF) which operates during the Contention-Free Period (CFP). This mode can be used only in an infrastructure network and uses the Point Coordinator (PC) as the network coordinator or arbiter. In a practical implementation, the PC can be included in the AP. Using a centralized polling based approach, the PC determines which station has the right to transmit and polls this station. A station can transmit only when polled by the PC. When polled, the station can transmit only one data frame which can be either a new data packet or a retransmission packet of a previous packet for which no ACK has been received. The station is allowed, though, to piggyback an ACK on the transmitted data frame. Thus, the PCF provides a contention-free data transfer.

To gain access to the medium and start the PCF period (in CFP), the PCF mode uses the PIFS interval (which is shorter in duration than the DIFS interval) along with special management frames that include the duration of the network allocation vector (NAV) to be set by all other stations. Using PIFS, the PC has the highest channel access priority. The PC sends the NAV value in the Beacon message. All stations set their NAV values which prevent them from accessing the channel during the CFP, therefore preventing collisions.

Thus, PCF provides collision-free access to the channel which has the potential to provide QoS support. As discussed later in this chapter, the standard does not provide the polling intelligence in terms of when to poll each station. It is left for the implementer to decide the polling intelligence.

4.1.4 Physical Layer Convergence Protocol (PLCP)

The PLCP layer located beneath the MAC has several responsibilities. It carries out the Carrier Sensing (CS) of the wireless media to determine whether there are ongoing transmissions on the channel. It also synchronizes the speed of the communicating stations' transmitter and receiver and establishes the communication. To accomplish these responsibilities it adds additional information to each packet.

The CS mechanism can be configured with several Clear Channel Assessment (CCA) options:

- *CCA Mode 1:* Energy above threshold. CCA reports a busy medium upon detecting any energy above the threshold.
- *CCA Mode 4:* Carrier sense with timer. CCA starts a timer whose duration is defined in the standard and reports a busy medium only upon the detection of a signal until the timer expires. Otherwise it reports the media is not busy.
- *CCA Mode 5:* A combination of carrier sense and energy above threshold.

The synchronization is obtained by using packet headers which are transmitted in increasing speeds. This means that the first part of the header is transmitted in a low speed, the second in a higher speed, and so forth. This mechanism allows speed synchronization but also slows down higher speed radios by as much as 30%.

4.1.5 QoS Support

As we mentioned previously, IEEE 802.11 MAC has two modes of operations, PCF and DCF, which deliver different QoS support. In order to clearly understand the QoS support, we first examine the QoS mechanisms provided by IEEE 802.11.

4.1.5.1 QoS in IEEE 802.11 DCF

QoS mechanisms (Figure 4.10)

1. *Classification:* There is no classification in DCF. Therefore, no service differentiation is provided.
2. *Channel access:* DCF uses contention-based media access control protocol (CSMA/CA).
3. *Packet scheduling:* The intrastation packet scheduling uses FIFO mechanism (i.e., all traffic in a station is queued and transmitted in a first-in-first-out order).

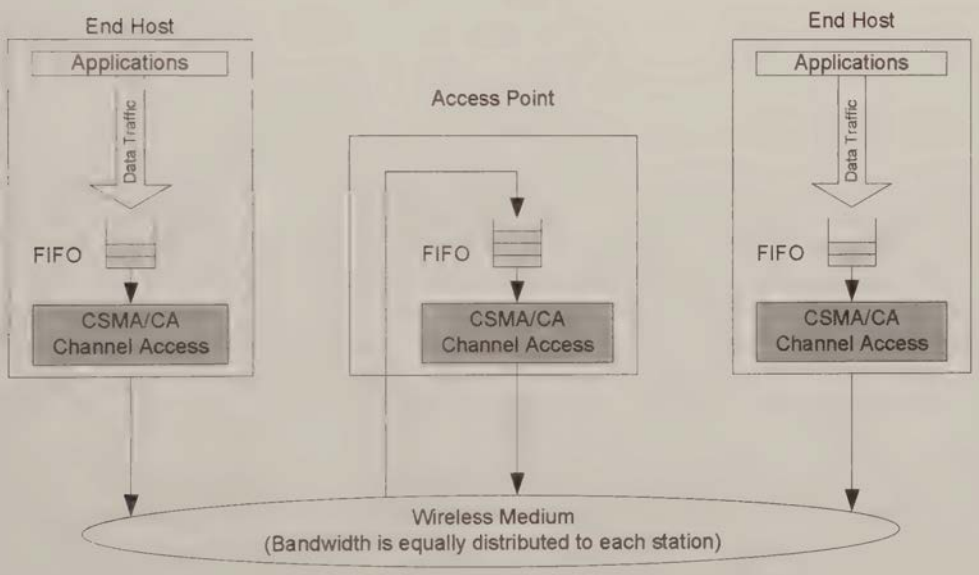


Figure 4.10 IEEE 802.11 DCF QoS Mechanisms

QoS service and supported applications

DCF delivers best effort QoS service level which is suitable for non-real time applications. There is no service differentiation and no service guarantee in terms of bandwidth and delay.

4.1.5.2 QoS in IEEE 802.11 PCF

QoS Mechanisms (Figure 4.11)

1. *Classification:* There is no classification in PCF. Therefore, there is no service differentiation within stations (i.e., all traffic in the same station is treated equally).
2. *Channel access:* PCF uses a polling based media access control protocol. A station is allowed to transmit packets only when it receives a polling message from the PC. Since no collisions occur, PCF can deliver a predictable service performance.
3. *Packet scheduling:* Due to the lack of a classification mechanism within each station, the intrastation packet scheduling uses a FIFO mechanism. The interstation packet scheduling is directly related to the polling sequence and polling frequency (the number of polls to a station in a time period) which are not specified by the standard.

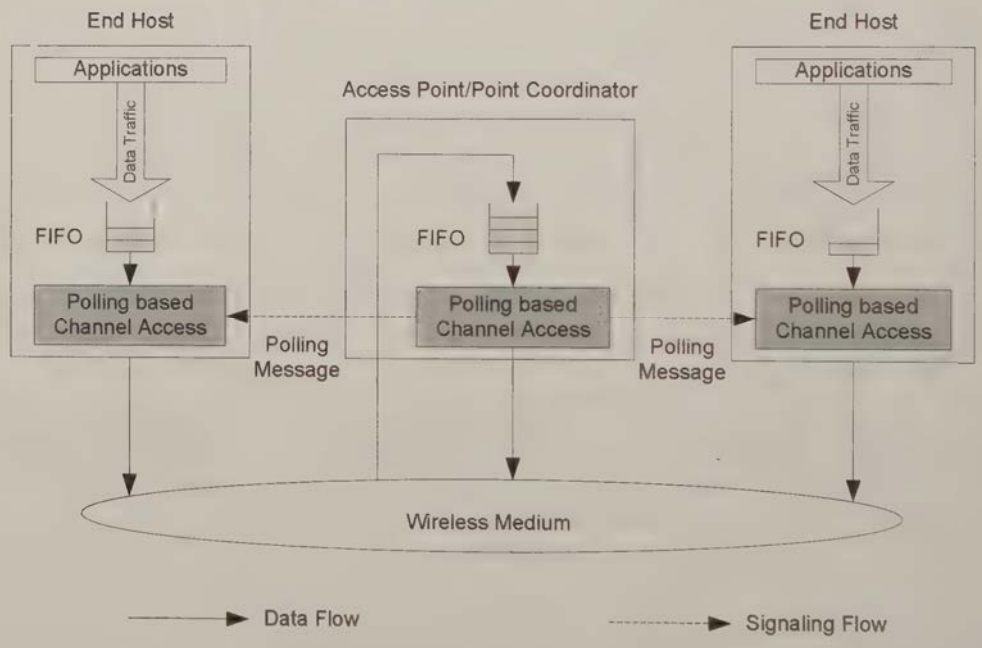


Figure 4.11 IEEE 802.11 PCF QoS Mechanisms

QoS service and supported applications

PCF can deliver a certain level of guaranteed QoS service which is suitable for real-time applications. To achieve guaranteed QoS service, the following functions have to be explicitly determined:

- Algorithms to determine the polling sequence and polling frequency to properly allocate bandwidth and satisfy applications' delay requirements.
- Algorithms to determine the duration of the CFP repetition interval, CFP interval, and CP interval so that applications' delay requirements are satisfied.
- QoS signaling: to determine the proper polling sequence and polling frequency, the PC needs to obtain from the stations information on the traffic dynamics at each station as well as applications' quantitative QoS requirements. Certain QoS signaling is included in IEEE 802.11e.

4.2 IEEE 802.11e (QoS Extension)

Realizing the shortcomings of the QoS support within the current IEEE 802.11 MAC, the IEEE organization (IEEE 802.11 Task Group E) has been involved in an extensive effort to specify MAC enhancements for better QoS support. This effort and potential extension to the standard are referred to as IEEE 802.11e. This extension has not been approved yet. Hence this discussion provides just a simplified explanation and glimpse into the proposals and ideas discussed within this effort. The proposed IEEE 802.11e applies to the physical layers defined in IEEE 802.11 and IEEE 802.11a, b, and g. In addition, the proposal allows for QoS support even when legacy IEEE 802.11 stations are present. This is an important feature that allows for friendly coexistence of new IEEE 802.11e compliant stations with legacy IEEE 802.11 stations.

Several problems with the PCF motivated this effort. Such problems include the unpredictable delays associated with beacons due to the CSMA/CA and PIFS mechanisms, unknown transmission durations of the polled stations caused by unknown frame lengths and modulation techniques, and the inability to differentiate among traffic streams which can enable the provision of higher priorities to QoS sensitive applications. These problems soften the QoS support in the following ways: 1) time delay is unbounded causing delay sensitive transmissions to be queued for a lengthy time, and 2) there is an inability to provide the needed bandwidth guarantees for bandwidth sensitive applications.

IEEE 802.11e's goal is to deliver the following:

- Quantitative QoS services (parameterized QoS): It provides QoS services that meet the application's quantitative requirements in terms of traffic specifications (TSPEC).
- Differentiated services (prioritized QoS): It provides priority services among traffic streams.

IEEE 802.11e proposes the following QoS enhancements:

- Traffic classification: traffic category and traffic stream
- Channel access and packet scheduling: enhanced MAC functions such as Enhanced DCF (EDCF) and Hybrid Coordination Function (HCF)
- QoS signaling
- New frames for QoS support

4.2.1 Traffic Classification

802.11e provides two types of classification: traffic category (TC) and traffic stream (TS).

4.2.1.1 Traffic Category

A traffic category provides a tool for applications to set a distinct priority relative to other data to be transmitted over the wireless media (prioritized QoS). 802.11e defines traffic categories to support differentiated services to at most eight delivery priorities designated 0 through 7. Data streams are classified to one of the eight traffic categories. By default, user priority TC7 is mapped to the highest delivery priority and user priority TC1 is mapped to the lowest delivery priority (see Table 4.2). It is possible to have multiple data streams mapped to the same traffic category. Therefore, classification based on traffic category is comparable to per-class classification defined in Chapter 3. Traffic Category cooperates with EDCF.

Table 4.2 IEEE 802.11e Traffic Categories

Priority Level	User Priority	Traffic Categories	Acronym	Traffic Type
Lowest	1	TC1	BK	Background
	2	TC2	--	Spare
	0 (default)	TC0	BE	Best effort
	3	TC3	EE	Excellent effort
	4	TC4	CL	Controlled load
	5	TC5	VI	Video <100 msec delay and jitter
	6	TC6	VO	Voice < 10msec delay and jitter
Highest	7	TC7	NC	Network control

To manage these traffic categories, IEEE 802.11e has to implement a number of physical queues. Table 4.3 shows a mapping between the traffic categories and the queues.

Table 4.3 IEEE 802.11e Mapping of Traffic Category to Queues

Number of Queues	Defining Traffic Type							
1	BE							
2	BE				VO			
3	BE				CL		VO	
4	BK		BE		CL		VO	
5	BK		BE		CL	VI	VO	
6	BK		BE	EE	CL	VI	VO	
7	BK		BE	EE	CL	VI	VO	NC
8	BK	--	BE	EE	CL	VI	VO	NC

4.2.1.2 Traffic Stream

To support quantitative QoS services (parameterized QoS), IEEE 802.11e defines traffic streams (TSs). In each station there are at most eight traffic streams. Each traffic stream is associated with its traffic specification (TSPEC) which includes:

1. Quantitative objectives for traffic attributes such as packet size and arrival rates
2. Traffic characteristics (constant vs. variable data rate, maximum delivery delay, maximum delay variance, etc.)
3. Acknowledgment policy

Each data flow with quantitative QoS requirement is directly classified to a traffic stream. Classification based on traffic streams is comparable to per-flow classification defined in Chapter 3. Traffic category cooperates with HCF.

Figure 4.12 shows the classification (both traffic category and traffic stream) in a station.

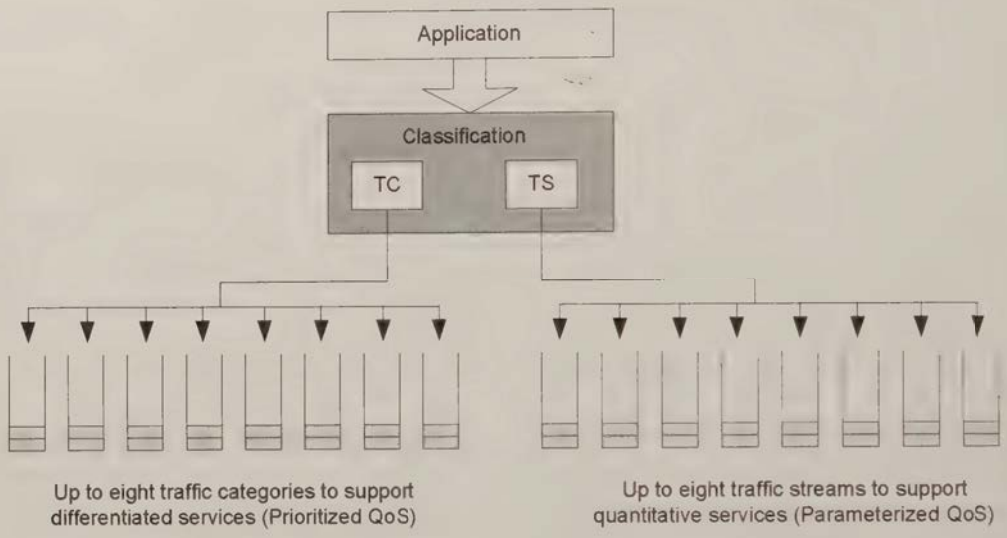


Figure 4.12 IEEE 802.11e Classification Mechanisms

4.2.2 Channel Access and Packet Scheduling

802.11e provides enhanced MAC functions (EDCF, HCF) which include both channel access and packet scheduling mechanisms. Therefore, we discuss both QoS mechanisms at once for each MAC.

IEEE 802.11e defines the following MAC protocol with two modes of operation: the Hybrid Coordination Function (HCF) to support QoS network configurations and Enhanced DCF (EDCF). Similar to 802.11, 802.11e proposal has two phases of MAC operation: Contention Period (CP) and Contention-Free Period (CFP), which alternate over time continuously. In each phase the station can send data packets or their fragments, according to rules set by each phase. The EDCF is used in the CP phase only, while the HCF is used in both phases (see Figure 4.13).

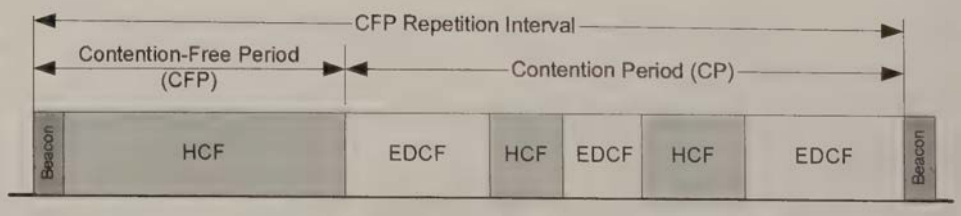


Figure 4.13 IEEE802.11e CFP Repetition Interval

The HCF combines functions from the legacy DCF and PCF with some enhanced, QoS-specific functions and frame subtypes. Such enhancements allow a uniform set of frame exchange sequences to be used for QoS transfers during both the CP and CFP. The HCF uses a contention based media access method, referred as enhanced DCF (EDCF). Stations may obtain transmission opportunities using one or both of these media access methods. Stations that support the 802.11e are referred as QoS Stations (QSTAs). The QoS station has an option to function as the centralized controller for all other stations within the WLAN or QoS supporting Basic Service Set (QBSS). This centralized controller is called Hybrid Coordinator (HC) or Point Coordinator (similar to 802.11 PCF). The HC is expected to be installed in the 802.11e Access Point (AP).

4.2.2.1 Enhanced Distributed Coordination Function (EDCF)

The EDCF provides differentiated, distributed access to the wireless medium for at most eight delivery priorities as defined above. EDCF channel access is derived from CSMA/CA used in DCF with the addition of priorities. Multiple packets are delivered through multiple backoff instances within one station. Each backoff instance is labeled with a TC specific parameter. In the CP phase, each TC within the stations contends for transmission and independently starts a backoff process congruent to the CSMA/CA after detecting the channel being idle for an Arbitration Interframe Space (AIFS). The AIFS is at least DIFS, and can be increased for each TC. Consequently, data streams with higher AIFS values, meaning longer backoff times, have lower priority access to the wireless media. Data streams with lower AIFS value will be able to access the wireless media before data streams with higher AIFS values (see Figure 4.14).

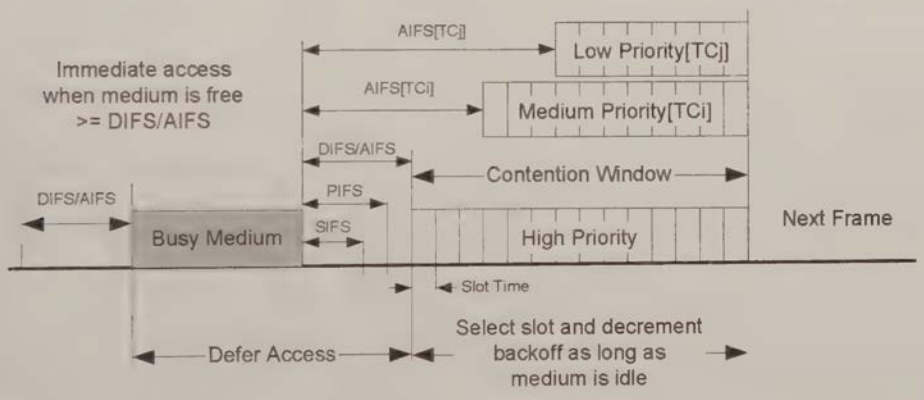


Figure 4.14 IEEE 802.11 EDCF

After the QoS stations wait for AIFS, each backoff sets a counter to a random number that is less than the size of the contention window (CW). Each traffic category has its own CW parameter. To achieve compatibility over legacy 802.11 WLANs, AIFS is set equal to DIFS. Similar to DCF, when the wireless medium is determined busy before the counter reaches zero, the backoff has to wait for the wireless medium to be idle for another AIFS period. If an unsuccessful

transmission occurred, a new enlarged CW is computed by using the persistence factor (PF). For example, if PF=1, CW remains unchanged. If PF=2, CW applies a binary exponential backoff algorithm. Therefore, the key parameters (maintained in each traffic category) that enable priorities are AIFS, CW, and PF.

The transmitting stations can retain control of the wireless medium by using the short interframe spaces (SIFSs) and the virtual carrier sense mechanism (NAV). The station can transmit a number of packets that will look like a single instance of activity to other contending stations.

A QoS station has the option to implement up to eight transmission queues based on TCs that dictate the queue priority. In case there are two or more TCs in the station and some of their backoff counters reach zero at the same time, the station's scheduler grants the transmission opportunity to the queue with the highest TC priority.

EDCF can still result in collisions among stations that carry traffic with the same priority. A station is allowed a maximum transmission duration (transmission opportunity [TXOP]) as defined in the Beacon message.

The RTS and CTS mechanism defined in 802.11 DCF can also be implemented in the EDCF.

QoS in EDCF

EDCF combines a collision based channel access (CSMA/CA) and priority packet scheduling to priority CSMA/CA (see Figure 4.15). Using traffic category classification and priority CSMA/CA, EDCF delivers differentiated QoS services (prioritized QoS).

4.2.2.2 Hybrid Coordination Function (HCF)

HCF defines more rules than EDCF allowing more control of transmissions (see Figure 4.16). The HCF allocates bandwidth and transmission opportunities (TXOPs) using a hybrid coordinator (HC) that has the highest access priority. HCF uses a centralized polling based approach similar to PCF. During CP, each TXOP starts as defined by the EDCF rules—that is, after AIFS plus backoff time, or when the station receives from the HC a special poll frame, referred as the QoS Contention-Free (CF)-Poll. The HC sends this QoS CF-Poll after it waits for the wireless media to be idle for PIFS. The QoS CF-Poll specifies TXOP, which is the time interval during which the station has the right to transmit. During the CFP, only the polled station is allowed to transmit. The CFP ends after the time announced in the beacon frame or by a CF-End frame from the HC.

In order to provide quantitative QoS services, HCF requires a signaling process that informs the HC about the transmission requirements of each traffic stream at each station. Using this information, the HC will determine which stations need to be polled, when, and which TXOP should be granted. The detailed signaling process will be discussed in the next section. TXOP is granted per-station—that is, the HC does not specify which traffic stream should be transmitted on the channel. It is up to the station to select the traffic stream to be transmitted.

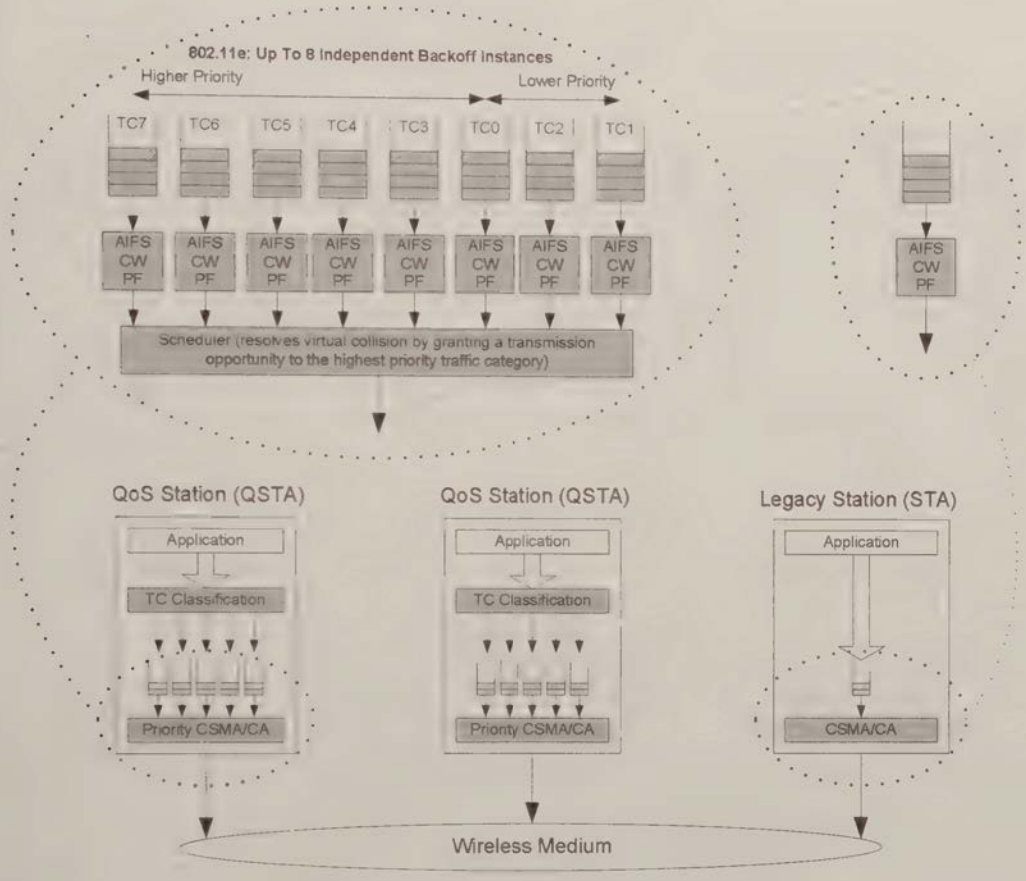


Figure 4.15 EDCF QoS Architecture

QoS in HCF

As shown in Figure 4.17, HCF contains the following QoS mechanisms: polling based channel access, QoS signaling, and traffic stream classification. These mechanisms enable the delivery of quantitative QoS services based on the applications' QoS requirements (TSPEC). However, IEEE 802.11e does not define the algorithms to set HCF parameters and it does not define the relationship between these parameters and the required QoS. Also, IEEE 802.11e does not present algorithms to compute the polling sequence and the time intervals provided in the HCF CF-Poll.

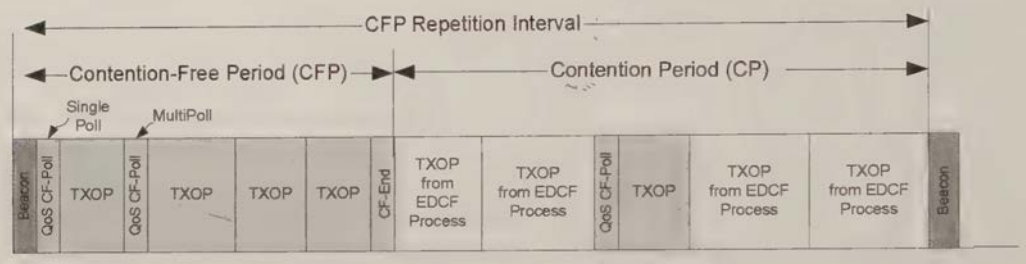


Figure 4.16 HCF CFP Repetition Interval

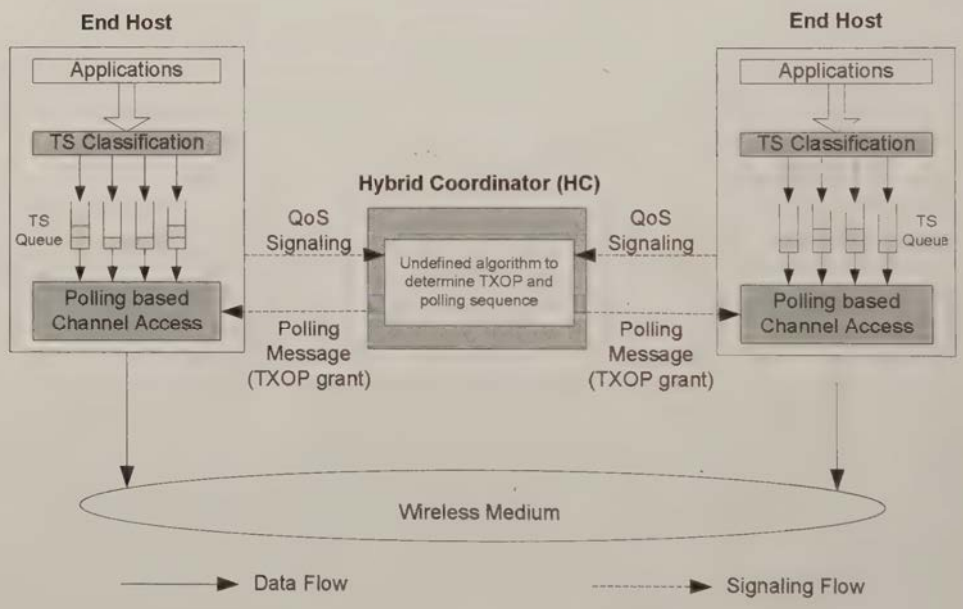


Figure 4.17 HCF QoS Architecture

4.2.3 QoS Signaling

The standard defines the QoS signaling which supports the operation of the QoS mechanisms. There are two forms of QoS signaling for traffic originating at the QoS stations: Queue State Indicator and Traffic Specification. QoS station transmits the message that contains Queue State or Traffic Specification to HC. QoS station can transmit the signaling message in three different ways.

1. *Transmit during CFP interval using HCF:* QoS station sends the signaling message during the CFP interval through the HCF polling process.
2. *Transmit during CP interval using EDCF:* The station sends the signaling message and contends with other stations for channel access.
3. *Transmit during CP interval using Controlled Contention (CC):* The station sends the signaling message during Controlled Contention Interval (CCI). The controlled contention process provides stations with the opportunity to request TXOPs by sending resource requests, without contending with other EDCF traffic. Each controlled contention instance starts when the HC sends a specific control frame that forces legacy IEEE 802.11 stations to set their NAV until the end of the controlled contention interval, thus they will refrain from transmission during the controlled contention interval (see Figure 4.18).

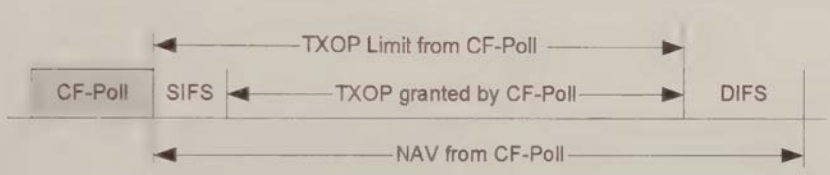


Figure 4.18 NAV Setting

4.2.3.1 Queue State Indicator Signaling

A QoS station transmits the queuing information of a traffic stream in the station to the HC. This will provide the information that is helpful to the HC to allocate the proper TXOP duration for the station to meet the traffic stream's demand. The algorithm that determines the TXOPs for each station using the queue information is not defined by the standard.

4.2.3.2 Traffic Specification Signaling

This signaling supports the resource reservation process in which a traffic stream in a QoS station transmits the signaling message that includes the traffic specification (i.e, bandwidth, delay requirement) to the HC. The signaling cooperates with the resource reservation process at higher layers (i.e., RSVP).



CHAPTER 5

HiperLAN

5.1 Introduction

The HiperLAN/2 set of standards was introduced by ETSI (the European Telecommunications Standards Institute) Project BRAN (Broadband Radio Access Networks). ETSI worked in collaboration with a HiperLAN2 Global Forum, which is a consortium of more than 50 companies established to market this new standard. In addition, HiperLAN/2 was developed in harmony with Japan's WLAN, referred to as HiSWANa. HiSWANa was developed by Japan's ARIB (Association of Radio Industries and Businesses) and its Multimedia Mobile Access Communications (MMAC) promotion association. Consequently, HiperLAN/2 and HiSWANa are very similar. Relatively few companies have products for HiperLAN/2 (e.g., Panasonic, and Amphion). HiperLAN/2 is designed to provide access to external IP, Ethernet, IEEE 1394, ATM, and 3G networks via an access point.

HiperLAN/2 provides data rates of up to 54 Mbps using Orthogonal Frequency Division Multiplexing (OFDM) radio technology for a range of up to 150 meters. This generation of standards supports both asynchronous data and applications that require QoS support. Notice that this standard, as other standards, defines signaling procedures required for the implementation of such QoS support. However, the algorithms that use this signaling information are not defined and are left for the developer.

The Media Access Control (MAC) protocol implements a form of dynamic time division duplex (TDD) and dynamic time division multiple access (TDMA) to provide connection oriented service and QoS support. The standard also supports station mobility with speeds of up to 10 ms.

The predecessor of HiperLAN/2, HiperLAN/1, was geared for ad hoc networking and asynchronous data. It applied the carrier-sense multiple access with collision avoidance (CSMA/CA) mechanism to resolve contention. HiperLAN/1 is considered a best effort system in spite of the fact that it included some preliminary mechanisms for QoS support.

5.2 Architecture

5.2.1 Network Topology

HiperLAN/2 distinguishes between a business environment and a residential environment. For a business environment the standard envisions a number of Access Points (APs) covering a certain area which may or may not overlap, depending on the business needs (Figure 5.1). This network topology is called the cellular access network configuration. An AP provider interconnects all mobile terminals (MTs) associated with it. All communication goes through the AP.

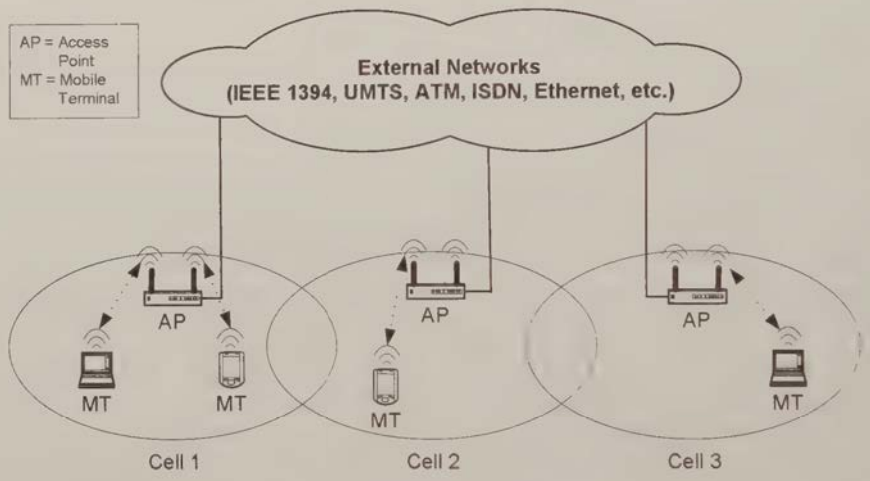


Figure 5.1 HiperLAN/2 Business Network (cellular access network configuration)

The residential environment is similar to the business environment. In addition, it can be operated in an ad hoc manner in which the MTs communicate with each other directly (see Figure 5.2).

The network topology for the residential environment is called an ad hoc LAN configuration. An ad hoc LAN configuration also requires a station called a central controller (CC), which provides network configuration control to all devices within the subnet. A CC in an ad hoc LAN configuration is similar to the AP in the cellular access network configuration. Unlike the dedicated hardware required for the AP, a CC is dynamically selected from the HiperLAN/2 mobile devices. When an MT which acts as a CC leaves the network, the CC responsibility can be handed over to another MT. HiperLAN/2 allows for multiple subnets and their respective CCs in the residence environment, similar to multiple cells and their respective APs in the business environment. Such multiple subnets can coexist since they operate at different frequencies.

HiperLAN/2 allows for two modes of operation: Centralized mode and Direct mode. In Centralized mode all traffic has to pass through the AP that manages the access to the wireless

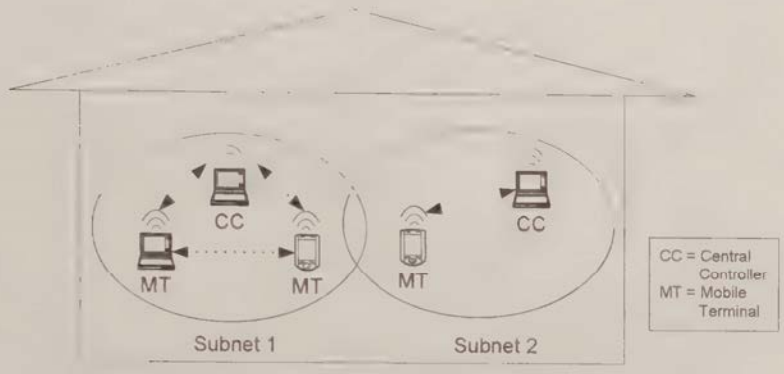


Figure 5.2 HiperLAN/2 Residential Network (ad hoc network configuration)

media. This traffic includes both traffic between mobile devices in the network and traffic between the mobile device and the outside network. In Direct mode the wireless media access is still managed by the CC. However, the traffic between the mobile devices is exchanged directly without going through the CC. Since the CC can also be connected to a core network, it is required that the CC will be able to work in both Direct and Centralized modes.

5.2.2 Protocol Stack

As shown in Figure 5.3, HiperLAN/2 defines the following three layers: the Convergence layer, the Data Link Control (DLC) with its MAC and other functionality, and the Physical layer.

The convergence layer provides the protocol interface between the upper layer and the DLC layer. The standard envisions that HiperLAN/2 provides wireless access to the external or core networks such as Internet Protocol (IP) networks, Asynchronous Transfer Mode (ATM) networks, 3G networks, and networks that use IEEE 1394 (Firewire) protocols. The data transport function of the convergence layer provides message format transformation (i.e., segmentation and reassembly function) between the higher layer and the DLC layer.

The data link layer contains three main entities: the Radio Link Control (RLC), the Error Control (EC), and Media Access Control (MAC). The user data plane function of the DLC receives the data packets from the upper layer (i.e., CL layer), provides the error control mechanism, and delivers the DLC data packets through the MAC. The RLC which is the control plane function of the DLC provides the Radio Resource Control (RRC), Association Control Function (ACF), and DLC Connection Control (DCC).

The HiperLAN/2 MAC of the AP/CC controls all transmissions over the wireless media. This includes uplink traffic from the station to the AP/CC and downlink traffic to the station. The MAC uses time division duplex (TDD) and dynamic time division multiple access (TDMA).

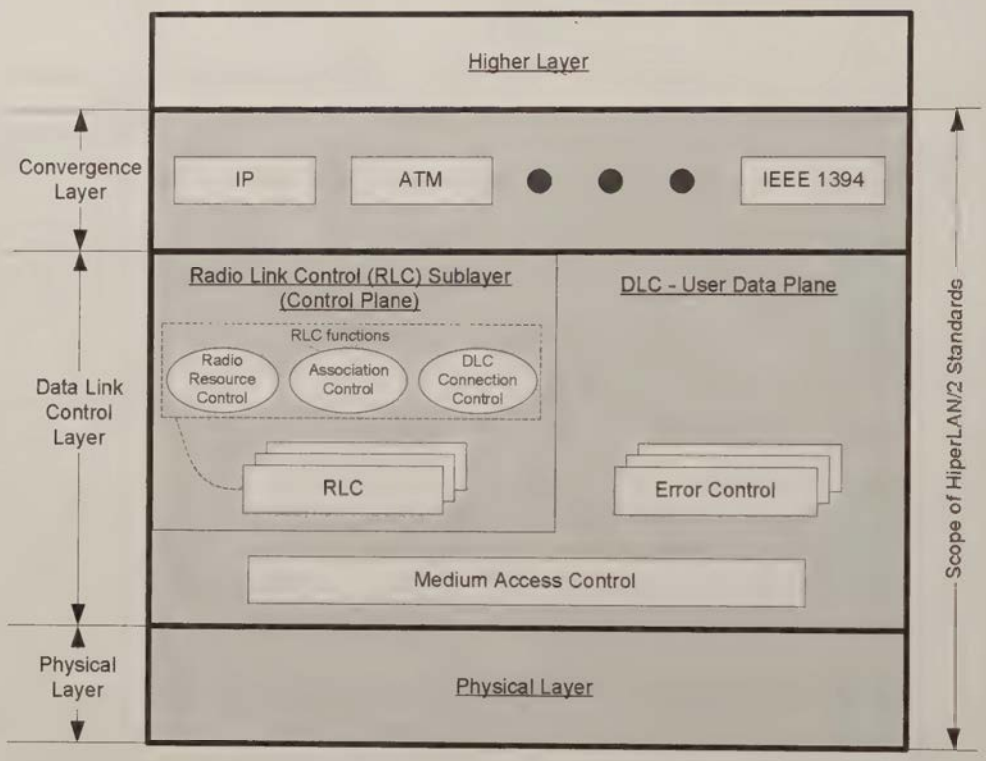


Figure 5.3 HiperLAN/2 Protocol Stack

5.3 Physical Layer

The physical layer supports several modes of transmissions: 6, 9, 12, 18, 27, 36, and 54 Mbps. Each mode employs a different radio modulation technique as shown in Table 5.1.

Table 5.1 HiperLAN/2 Radio Modulation Techniques

Modulation	Code Rate	Physical Layer Bit Rate (Mbps)
BPSK	1/2	6
BPSK	3/4	9
QPSK	1/2	12

Table 5.1 HiperLAN/2 Radio Modulation Techniques (Continued)

Modulation	Code Rate	Physical Layer Bit Rate (Mbps)
QPSK	3/4	18
16QAM	9/16	27
16QAM	3/4	36
64QAM (optional)	3/4	54

HiperLAN/2 operates at 5 GHz frequency spectrum as shown in Figure 5.4. In the U.S., 300 MHz bandwidth in U-NII band has been allocated, in Japan 100 MHz (with sharing rule) bandwidth has been allocated, and in Europe 455 MHz bandwidth has been allocated in license exempt band. The channel spacing is 20 MHz, which allows high bit rates per channel but still has a reasonable number of channels in the allocated spectrum (e.g., 19 channels in Europe).

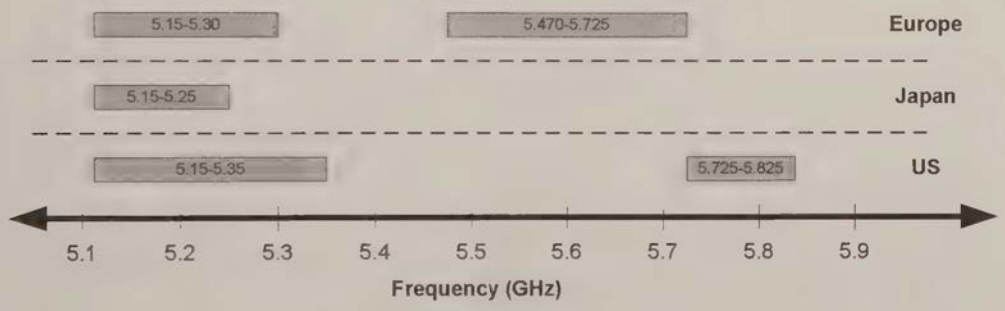


Figure 5.4 HiperLAN/2 Frequency Spectrum Allocation

HiperLAN/2 uses orthogonal frequency division multiplexing (OFDM). The transmission unit is a burst, which consists of a preamble part and a data part, where the latter could originate from each of the transport channels within the DLC.

HiperLAN/2 defines Dynamic Frequency Selection (DFS), which allows sharing of the wireless media among several HiperLAN/2 networks. The AP chooses frequencies based on a frequency selection process considering interference measurements. In addition, the AP/CC will adapt its transmission mode (speed, code, and modulation) based on its measurements of the transmission link quality. Moreover, the mobile stations, based on their measurements, may suggest the preferred transmission mode.

5.4 Data Link Control (DLC) Layer

As mentioned earlier, the DLC layer contains three sublayers: Error Control (EC), Radio Link Control (RLC), and Media Access Control (MAC). In addition, there are logical and transport channels as shown in Figure 5.5. Logical channels are identified by the type of message they carry while transport channels are identified by the message format and the channel access method. The control messages and user data messages (which originate from the DLC or higher layer) are mapped to the appropriate logical channel based on their contents. Then the messages are passed to the transport channel in order to construct the MAC frame and receive the appropriate channel access. In the next subsections we introduce the functionality of each DLC sublayer, the MAC frame, and the transport and logical channels. Then we describe MAC protocol examples.

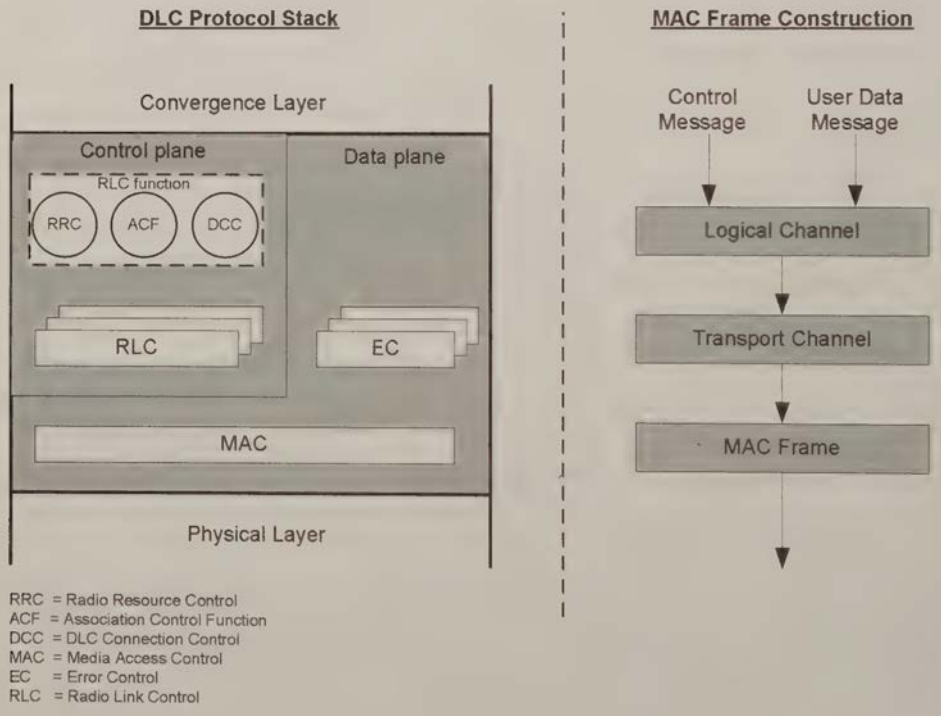


Figure 5.5 DLC Protocol Stack and MAC Frame Construction

5.4.1 Radio Link Control (RLC)

The RLC manages the network and thus exchanges control data between the AP and the MTs. The AP can have multiple RLC instances, where each RLC instance associates with an MT. There is only one RLC instance in an MT.

The RLC has three functions: Radio Resource Control (RRC), Association Control Function (ACF), and DLC Connection Control (DCC).

The RRC is responsible for detecting and efficiently using the available radio resources. It manages handover, dynamic frequency selection, station alive/absent, power saving, and power control. The RRC selects the frequency range on which the communication will be conducted and, when needed, it decides to move to a different frequency range. This is done based on the AP/CC own channel measurements as well as the mobile station channel measurements. To preserve the battery power of the stations, the RRC defines when and for how long to put the MTs into sleep mode. The lengths of sleep intervals are negotiated between the AP/CC and the MTs.

The ACF manages the association process for authentication, key management, association, disassociation, and encryption. A station that wants to communicate over the wireless media is required to be associated with an AP/CC. This is due to the fact that the AP/CC allocates wireless resources for each associated station. Moreover, the MAC is centrally controlled by the AP/CC in both Centralized Mode and Direct Mode. When a new MT joins the network by performing the association process with the AP's ACF, the ACF assigns a unique MAC Identification number (MAC ID) to the MT. The ACF also maintains information of the MT's capabilities in terms of whether encryption and authentication are performed, and which encryption and authentication methods are employed.

The DCC sets up, maintains, and releases user connections. If any kind of QoS support is requested, the connection has to supply the QoS parameters. The specification of these parameters is not part of the HiperLAN/2 standard. The connection setup includes procedures for centralized mode or direct mode, procedures for connection release, and procedures for joining and leaving multicast groups.

5.4.2 Error Control (EC)

The EC is responsible for detection and recovery from transmission errors on the wireless media. It also ensures in-sequence delivery of packets. Each connection is supported by an instance of the EC.

The EC is based on an Automatic Repeat Request (ARQ) algorithm. Additional error correction techniques can be employed. The ARQ scheme is based on a selective repeat mechanism in which the receiver requests that the sender retransmit packets determined to be erroneous. In addition, the EC module provides several modes to improve the transmission reliability:

1. *Acknowledged mode:* The EC will retransmit the acknowledgments from the receiver. Low latency can be maintained by a mechanism that discards packets based on time information.

2. *Repetition mode:* The EC will repeat the transmission. No acknowledgments are available in this mode. Thus, this mode is typically used for broadcast data.
3. *Unacknowledged mode:* This is for unreliable, low latency transmission without retransmissions.

Therefore, unicast data can be sent using either the Acknowledged mode or the Unacknowledged mode. Broadcast data can be sent using either the Repetition mode or the Unacknowledged mode. Multicast data can be sent in the Unacknowledged mode or can be piggybacked into other unicast transmissions.

5.4.3 Media Access Control (MAC)

The HiperLAN/2 MAC that resides at the AP/CC controls all the transmissions over the wireless media. This includes 1) uplink transmissions from the MTs to the AP/CC, 2) downlink transmission from the AP/CC to the MTs, and 3) direct transmission among the MTs. Direct transmission is mandatory for the residential or ad hoc configuration. Therefore, the MAC deploys time division duplex (TDD) and dynamic time division multiple access (TDMA). The channel is structured into MAC frames as shown in Figure 5.6. The MAC frame has a fixed duration of 2 ms. Each MAC frame starts with the Broadcast Channel (BCH) duration.

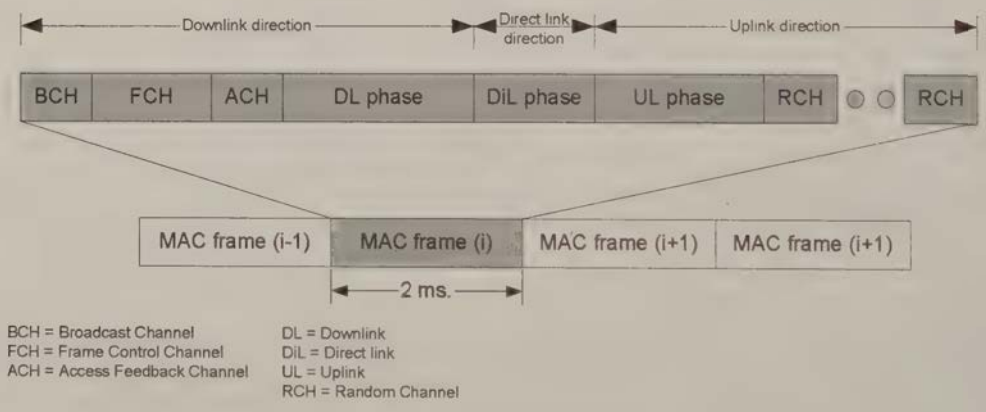


Figure 5.6 MAC Frame Structure

As shown in Figure 5.7, the frame structure slightly differs if the AP/CC has a sectored antenna. In this case, each phase is repeated in time, one for each sector. The use of Direct Link (DiL) with sectored antennas is not specified.

The MAC frame consists of: Broadcast Channel (BCH) duration, Frame Control Channel (FCH) duration, Access Feedback Channel (ACH) duration and at least one Random Channel (RCH) duration. If there is transmission between the AP/CC and the MTs, the Downlink (DL)

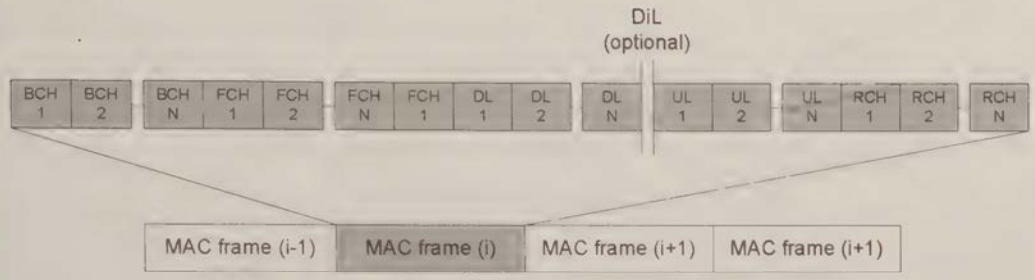


Figure 5.7 MAC Frame Structure for Sectored Antennas

phase and/or Uplink (UL) phase are included in the MAC frame. If there is transmission among the MTs (direct mode), the Direct Link (DiL) phase is also included. The BCH duration is fixed. The duration of the FCH, DL, DiL, and UL phases and the number of RCHs are dynamically determined by the AP/CC.

The BCH, FCH, ACH, and RCH contain control messages, whereas the DL phase, DiL phase, and UL phase mostly contain user data and certain control messages. The standard defines the message format through the transport and logical channels. There are six transport channels (with three letter abbreviations):

1. Broadcast Channel (BCH)
2. Frame Channel (FCH)
3. Access Feedback Channel (ACH)
4. Long Transport Channel (LCH)
5. Short Transport Channel (SCH)
6. Random Access Channel (RCH)

There are ten logical channels (with four letter abbreviations):

1. Broadcast Control Channel (BCCH)
2. Frame Control Channel (FCCH)
3. Random Access Feedback Channel (RFCH)
4. User Data Channel (UDCH)
5. User Multicast Channel (UMCH)
6. User Broadcast Channel (UBCH)
7. RLC Broadcast Channel (RBCH)
8. Dedicated Control Channel (DCCH)
9. Link Control Channel (LCCH)
10. Association Control Channel (ASCH)

Figure 5.8 illustrates the MAC frame structure, the transport and logical channels and their relationships. All messages are mapped to specific logical channels based on their content and their functionality. The messages are constructed according to a format defined by the logical channel. Then the messages will pass to the transport channel to receive the appropriate transport services.

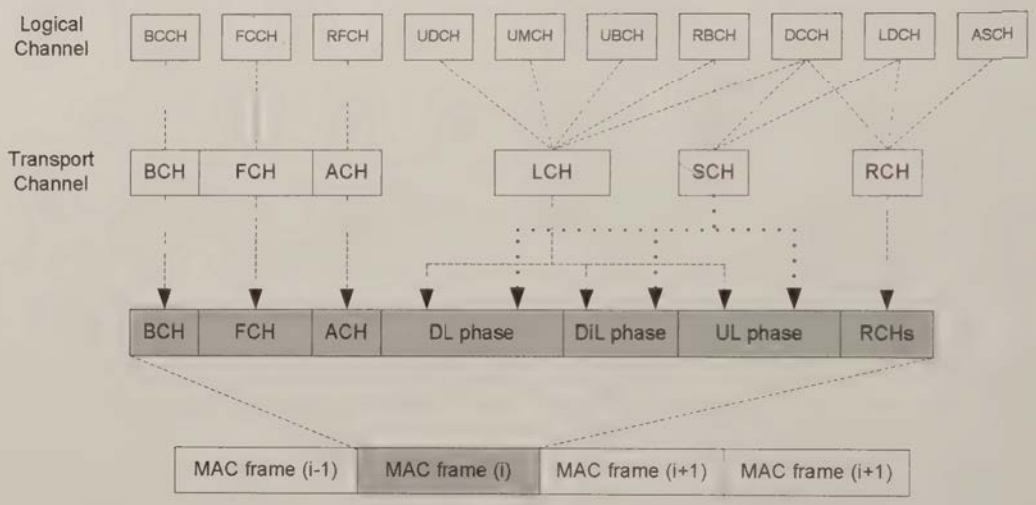


Figure 5.8 MAC Frame Structure, Transport Channel, and Logical Channel

Each MT's MAC is uniquely identified by a MAC ID which was assigned by the AP's RLC during the association process. The HiperLAN/2 MAC is connection oriented. Before an application can transmit, it requires the MT to establish a connection with the AP's RLC. This RLC assigns to this connection a DLC Connection ID (DLCC ID). In the centralized mode (CM), the DLC User Connection ID (DUC ID), which is obtained as a combination between the MAC ID and the DLCC ID, uniquely identifies the connection in the cell. In distribution mode (DM), the definition of DUC ID is slightly different. DUC ID, which identifies the connection in DM mode, is the combination of source MAC ID, destination MAC ID, and DLCC ID. DUC ID is also used for identifying the control channels. DUC ID is essential for the AP to allocate bandwidth to a specific connection (both user data and control).

5.4.4 Logical Channel

A logical channel is viewed as a generic term for distinct data paths. Logical channels are defined by the type of information they carry.

5.4.4.1 Broadcast Control Channel (BCCH)

The BCCH is used in the downlink direction and carries BCCH information about the entire wireless cell to all the stations within the cell. The BCCH message is only forwarded to BCH transport channel and eventually is transmitted on the BCH. BCCH contains: Net ID (the network or the cell identifier), AP ID (AP identifier), AP TX level (indicating the transmission power of the AP), AP RX UL Level (indicating the expected reception power level of the AP). Moreover, the BCCH also contains a pointer or a location and the length of the FCH duration and RCH duration in the MAC frame as shown in Figure 5.9. The support provided by BCCH is mandatory for APs and CCs. The mobile stations should be able to interpret the BCCH.

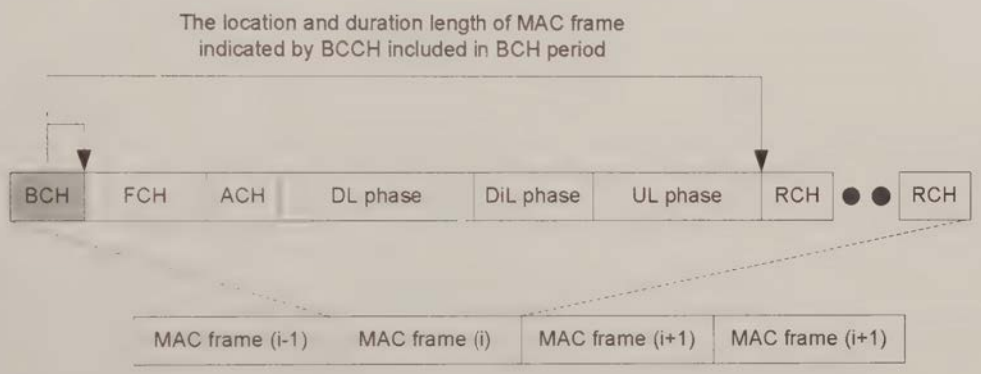


Figure 5.9 The Pointers in BCCH

5.4.4.2 Frame Control Channel (FCCH)

The FCCH is used in the downlink direction and carries information that describes the structure of the MAC frame visible at the air interface (see Figure 5.10). FCCH is also called Resource Grants (RGs). An RG, which is per connection basis (DUC ID), corresponds to a number of transport channels (the number of LCHs or SCHs), the PHY modes to be used, and the location in the frame where the reception/transmission will take place. The connection can be a control connection (RBCH, DCCH, LCCH) or a user data connection (UDCH, UBCH, UMCH). In addition, FCCH includes announcements of the empty parts in the MAC frame. FCCH becomes an Information Element (IE) of the FCH transport channel. The FCH can contain multiple IEs. Therefore, the FCH's duration is variable. The support of the FCCH is mandatory for APs and CCs. MTs are required to be able to interpret the FCCH.

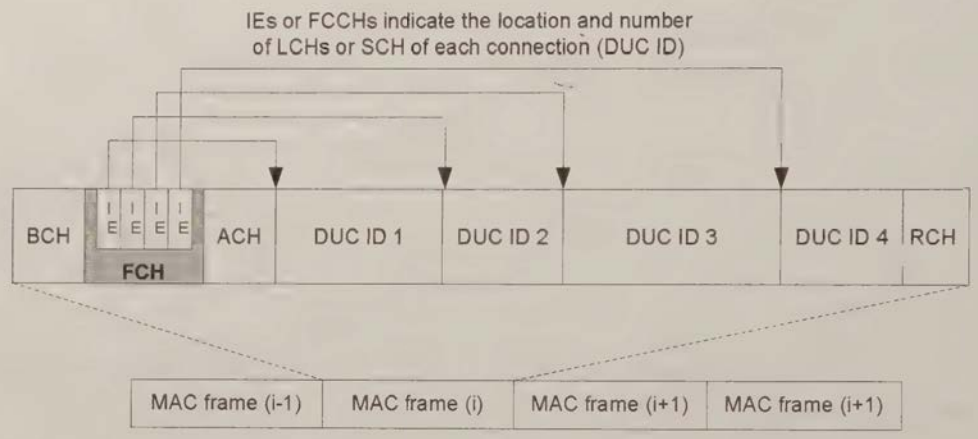


Figure 5.10 Frame Control Channel (FCCH) in FCH

5.4.4.3 Random Access Feedback Channel (RFCH)

The RFCH carries information for stations that have used the RCH in the previous MAC frame. This information contains the result of their access attempts. The RFCH message, which is transmitted once per MAC frame, is included in the ACH (see Figure 5.11). The support of the RFCH is mandatory for APs and CCs. MTs are required to be able to interpret the RFCH.

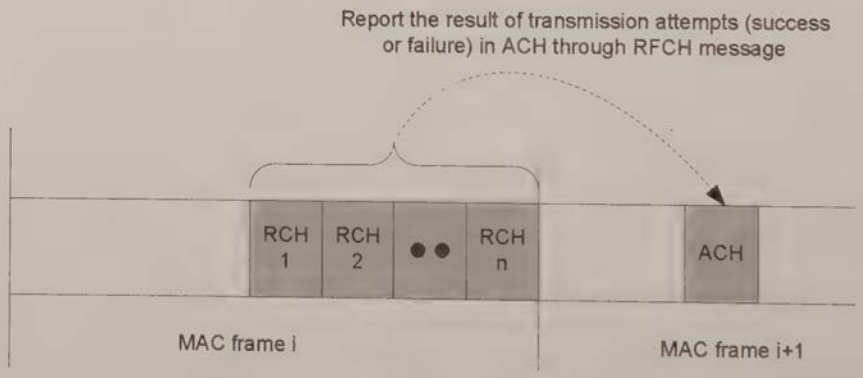


Figure 5.11 ACH Operation

5.4.4.4 RLC Broadcast Channel (RBCH)

The RBCH is used in the downlink direction in Centralized Mode (CM) and may be used by an MT when it communicates without the AP/CC in Direct Mode (DM). RBCH carries broadcast control information about the whole radio cell. RBCH transmission is triggered by the AP/CC in

the centralized mode and is determined by the originating MT in the direct mode. The following information may be included in the downlink RBCH: broadcast RLC messages, transmission of the assigned MAC ID to a non-associated mobile terminal, convergence layer ID information, and encryption related information. Figure 5.12 shows the RBCH frame construction process. The support of the downlink RBCH in AP/CC and MTs is mandatory.

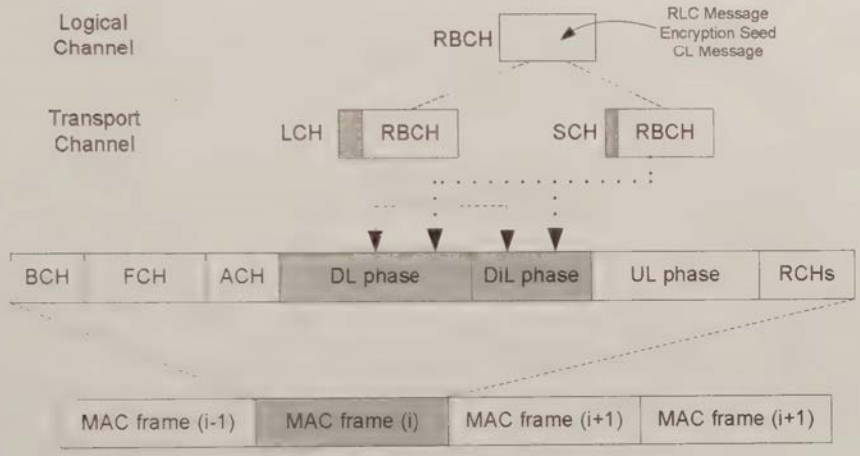


Figure 5.12 RBCH Frame Construction Process

5.4.4.5 Dedicated Control Channel (DCCH)

The DCCH is used to carry RLC messages. It is used in the uplink, downlink, and Direct Mode communication. The DCCH may also be used for the transmission of Resource Requests (RRs) in the uplink direction. DCCH is transmitted through LCH or SCH or RCH transport channel. Figure 5.13 shows the DCCH frame construction process. The support of the DCCH in both MTs and AP/CC is mandatory. It is required in Direct Mode if this feature is supported.

5.4.4.6 User Broadcast Channel (UBCH)

The UBCH is used to transmit user broadcast data. The UBCH can be sent by the AP/CC in the downlink and by an MT in the direct link. The UBCH is transmitted in repetition or unacknowledged mode. UBCH is transmitted through LCH transport channel. UBCH frame construction process is shown in Figure 5.14. The support of the UBCH in both MTs and AP/CC is mandatory. It is required in Direct Mode if this feature is supported.

5.4.4.7 User Multicast Channel (UMCH)

The UMCH is used to transmit user multicast data. The UMCH can be sent by the AP/CC in the downlink and by an MT in direct link. The UMCH is transmitted in repetition or unacknowledged mode. UMCH is transmitted through LCH transport channel. UMCH frame construction process is shown in Figure 5.14. The support of the UMCH in both MTs and AP/CC is mandatory. It is required in Direct Mode if this feature is supported.

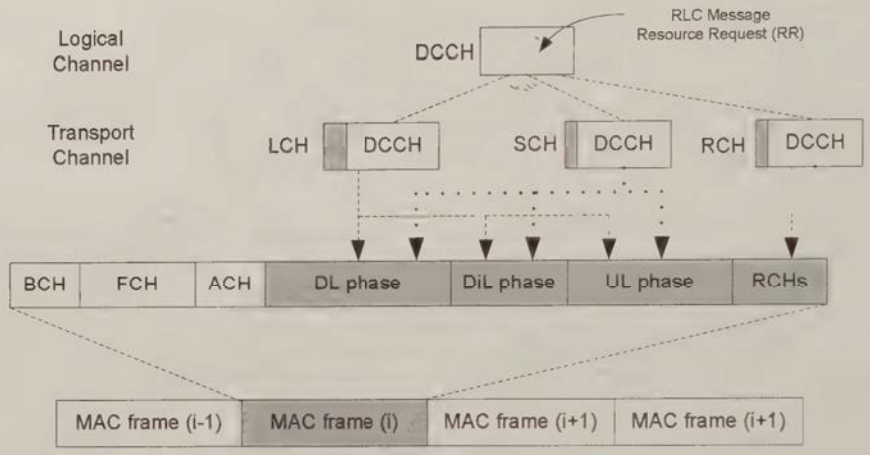


Figure 5.13 DCCH Frame Construction Process

5.4.4.8 User Data Channel (UDCH)

The UDCH is used to transmit user data between the AP and the MTs in Centralized Mode (CM) or between two MTs in Direct Mode (DM). A UDCH is always granted together with zero or more transport channels for a connection that is announced in the RG in the FCCH. UDCH is transmitted through LCH transport channel. UDCH frame construction process is shown in Figure 5.14. The support of UDCH for uplink and downlink is mandatory for both MTs and AP/CC. The support of the UDCH for Direct Mode is mandatory if this feature is supported.

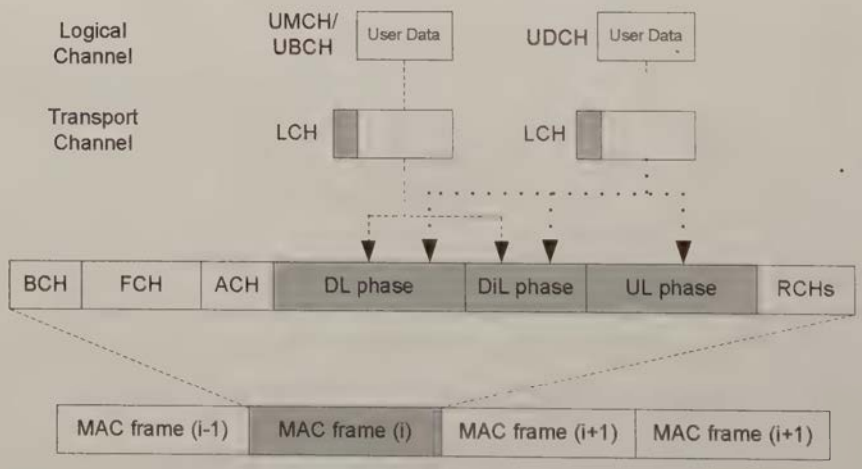


Figure 5.14 UMCH/UBCH/UDCH Frames Construction Process

5.4.4.9 Link Control Channel (LCCH)

The LCCH is used in both the uplink and downlink to transmit error control information such as ARQ feedback. It is used in either the Centralized or Direct Mode. The LCCH is also used in the uplink transmission of the RRs. LCCH is transmitted through the SCH or RCH transport channels. LCCH frame construction process is shown in Figure 5.15. The support of the LCCH for uplink and downlink is mandatory for both MTs and AP/CC. The support of the LCCH for Direct Mode is mandatory if this feature is supported.

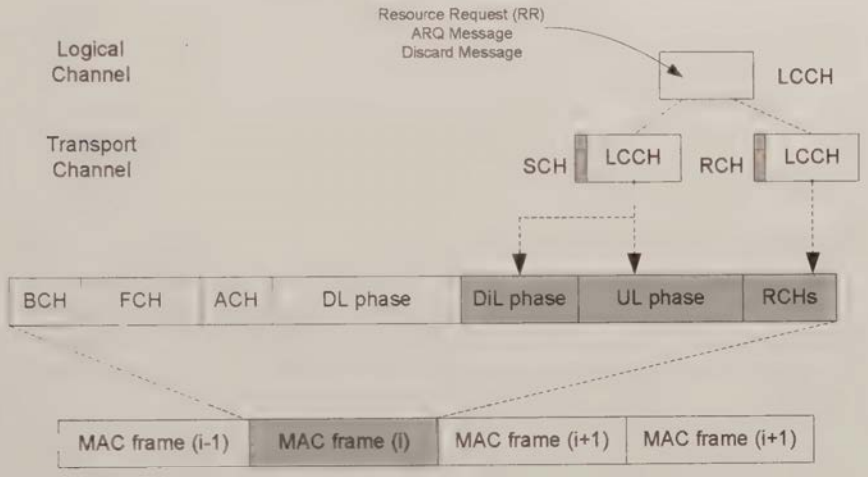


Figure 5.15 LCCH Frame Construction Process

5.4.4.10 Association Control Channel (ASCH)

The ASCH is only used in the uplink and carries new association and handover request messages. These messages are sent only by the MTs that are not associated to an AP/CC. ASCH is transmitted through the RCH transport channel. ASCH frame construction process is shown in Figure 5.16. The support of the ASCH is mandatory for AP/CC and MTs.

5.4.5 Transport Channels

The various logical channels are mapped into corresponding transport channels that describe the basic message format.

1. *Broadcast Channel (BCH):* The BCH carries the BCCH in the downlink direction. Its support is mandatory for AP/CC and stations.
2. *Frame Channel (FCH):* The FCH carries the FCCH. It is broadcasted in the downlink direction and is mandatory for all AP/CC and stations.
3. *Access Feedback Channel (ACH):* The ACH carries RFCH in the downlink. Its support is mandatory for AP/CC and stations.

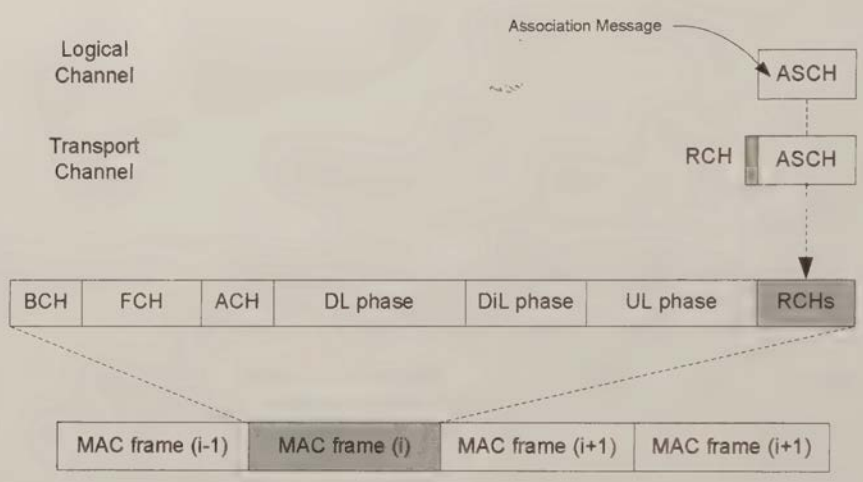


Figure 5.16 ASCH Frame Construction Process

4. *Long Transport Channel (LCH):* The LCH carries user data for the connections related to the granted UDCHs, UBCHs, and UMCHs, as well as control information for the connections related to the DCCH and RBCH. Its support is mandatory for AP/CC and stations.
5. *Short Transport Channel (SCH):* The SCH carries short control information for the DCCH, LCCH, and RBCH. Its support is mandatory for AP/CC and stations.
6. *Random Access Channel (RCH):* The RCH is defined for the purpose of giving a station the opportunity to send control information to the AP/CC when it has no granted SCH available. It can carry RRs, ASCH, and DCCH data. Its support is mandatory for AP/CC and stations.

5.4.6 Resource Request and Resource Grant

The standard defines the resource request (RR) and resource grant (RG) signaling mechanism to support the bandwidth allocation algorithm, which plays an important role in QoS support. The standard does not define the bandwidth allocation algorithm. Figure 5.17 shows a simplified diagram of the bandwidth allocation input and output.

Because HiperLAN/2 MAC is connection oriented, bandwidth allocation is performed per connection basis. Each RR and RG serves for an individual connection. In order for the AP/CC to allocate resources effectively it needs to know the bandwidth demand of each connection represented by the queue status (i.e., queue length). Thus, the stations will report their queue status in the RR messages to the AP/CC. Using the queue information received, the QoS parameters, and the available network resources, the bandwidth allocation algorithm which resides at the

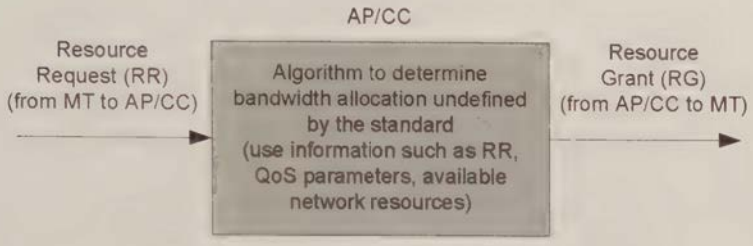


Figure 5.17 Bandwidth Allocation

AP/CC will determine the RG. The RG indicates when a connection is allowed to transmit and how many packets can be sent. The RG allocation algorithm which is not defined by the standard needs to be determined by the system developer.

The RR can operate in several ways:

1. *Polling Resource Request* (Figure 5.18): An MT initiates the RR by listening to the poll. First, the AP/CC indicates the polling for an RR by setting the RR Poll bit in the FCCH message. Then, the MT will transmit the RR using LCCH with SCH transport channel on the uplink.

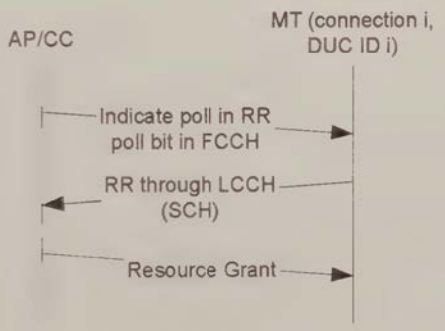


Figure 5.18 Polling Resource Request

2. *Unpolling Resource Request* (Figure 5.19): An MT initiates the RR, which is sent through DCCH with SCH transport channel (collision-free transport) and/or LCCH with RCH transport channel (collision-based transport). AP/CC can control the access delay of the RCH by changing the number of contention slots (i.e., the number of RCHs). If a collision occurs, the stations are informed in the ACH of the next MAC frame. Then, the station backs off a random number of access slots.

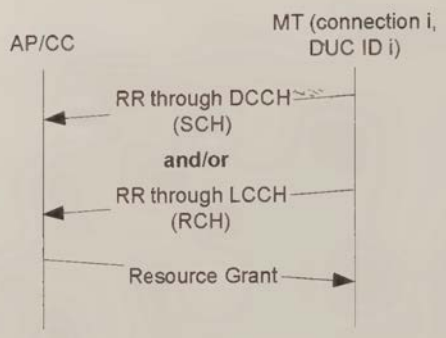


Figure 5.19 Unpolling Resource Request

5.5 Convergence Layer

The convergence layer (CL) maps service requests from the higher layers to the service offered by the DLC. It converts higher layer packets of fixed (ATM) or variable (Ethernet) length into fixed-length packets, referred to as Service Data Unit (SDU), that are used within HiperLAN/2. For example, CL will map the 802.1p priority field into a HiperLAN/2 priority class.

5.6 QoS support

HiperLAN/2 defines several QoS mechanisms to provide QoS support for multimedia applications. HiperLAN/2 manages the resources based on a connection-oriented approach which enables the finest granularity of QoS support (per flow and quantitative QoS services). In this section, we describe each QoS mechanism provided by HiperLAN/2.

5.6.1 Classification

Each connection is tagged with DUC ID. In centralized mode (CM), DUC ID consists of MAC ID and DLC connection ID (DLCC ID). In distributed mode (DM), DUC ID consists of source MAC ID, destination MAC ID and DLCC ID. DUC ID enables classification. Classification identifies packets based on DUC ID and forwards them to the appropriate queue. Therefore, HiperLAN is capable of per-flow classification which supports per-flow QoS services.

5.6.2 Channel Access

As described in Section 5.4.3, the channel access is TDD/TDMA, which is controlled by the AP/CC through FCCH. Since TDD/TDMA is a collision-free channel access scheme which provides tight channel access control, HiperLAN/2 can provide QoS support for applications with strict QoS requirements.

5.6.3 Packet Scheduling

The packet scheduling algorithm 1) allocates bandwidth for connections in terms of the number of packets (i.e., number of LCHs) and 2) determines when a connection is allowed to transmit. This information is defined in the FCH IEs (FCCH). The packet scheduling algorithm uses the RR and RG, which are described in Section 5.4.6. The standard does not define the packet scheduling algorithm. It just defines the signalling mechanism such as RR and RG.

Figure 5.20 summarizes the HiperLAN/2 QoS architecture which provides the necessary mechanisms to deliver per-flow quantitative QoS services.

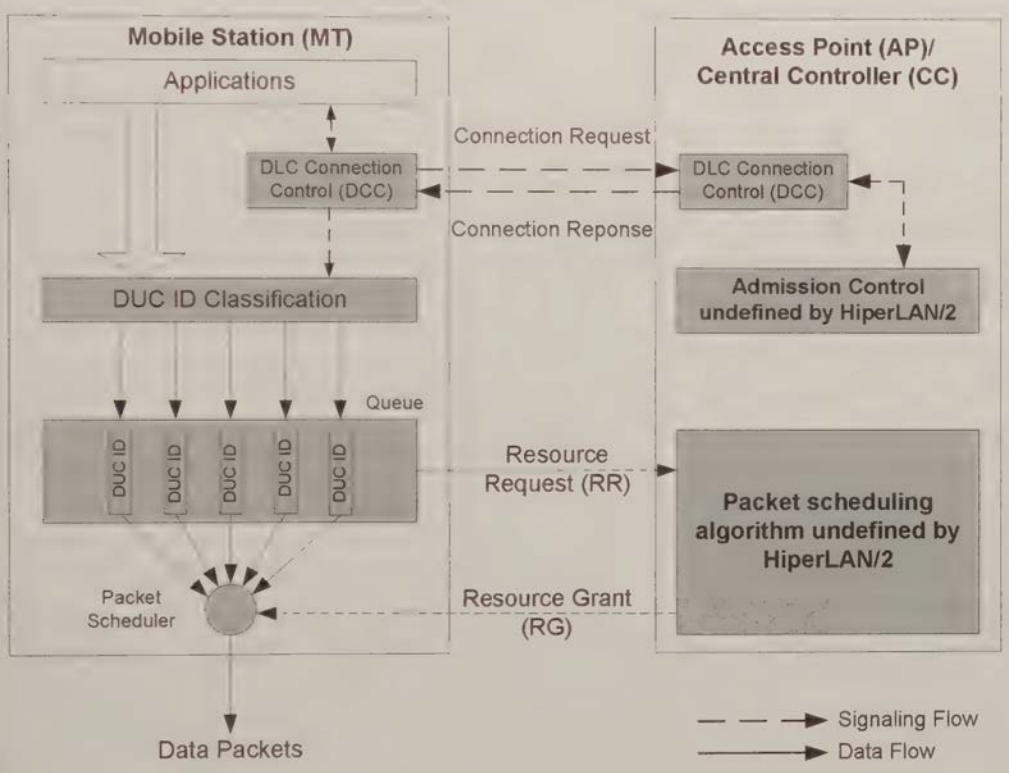


Figure 5.20 HiperLAN/2 QoS Architecture



CHAPTER 6

HomeRF

6.1 Introduction

HomeRF 2.0 standard was developed by the HomeRF Working Group (HRFWG) formed by companies such as Siemens, Proxim, National Semiconductor, Xilinx, Motorola, and AT&T. The group goal is to establish the mass deployment of interoperable wireless networking access devices to both local content and the Internet for voice, data, and streaming media in consumer environments. Relative to IEEE 802.11, few products have been announced.

Companies such as Cayman Systems, Compaq, Intel, Motorola, and Proxim have announced HomeRF 2.0 based communication cards. Utilizing the HomeRF QoS support for voice, companies such as Motorola and Siemens introduced products that include voice delivery. Motorola introduced the Simplefi system, which allows users to transmit Internet digital audio to the existing stereo equipment. Siemens introduced Voice Data Gateway and HomeRF based phones.

HomeRF 2.0 specification defines a common interface that supports wireless voice and data networking in the home. The specification allows 800 kbps rate for isochronous voice and rates of 1.6, 5, and 10 Mbps for asynchronous data transfer. HomeRF specification is backwards compatible with HomeRF 1.2, which operates at rates of 0.8 Mbps and 1.6 Mbps. HomeRF targets to support a wide range of applications by defining the following three types of data services:

1. *Asynchronous data service:* It provides best effort service for asynchronous data applications such as file transfer and email.
2. *Priority asynchronous data service:* It provides priority service for asynchronous data applications with loose QoS requirements, such as streaming media (video, audio).
3. *Isochronous data service:* It provides support for applications with strict QoS requirements (i.e., delay, delay jitter) such as interactive voice and cordless phone.

HomeRF 1.x standard did not describe roaming procedures in terms of device hand-offs and resolution of multiple address conflicts. HomeRF 2.0 added the possibility of an extended network that is an aggregation of individual wireless HomeRF networks associated with the same network ID number. Such a network is managed by roaming capable Connection Points (CPs). The standard also defines roaming procedures which include hand-offs, addresses, and data transfers between roaming devices.

6.2 Architecture

6.2.1 Network Topology

Before we describe the HomeRF network topology, we would like to introduce the HomeRF devices and their functionalities. The HomeRF standard categorizes the HomeRF devices into three main types, based on the type of application the devices support:

1. *Asynchronous data device* (also called A-node): These devices, such as PC, Laptop, and PDA, support asynchronous data services (i.e., file transfer, email).
2. *Streaming data device* (also called S-node): These devices, such as audio headsets, support priority asynchronous data services (i.e., audio or video streaming).
3. *Isochronous data device* (also called I-node): These devices, such as cordless phones, support isochronous data services (i.e., telephone call).

Some devices can support multiple services. For example, AI-nodes combine the functionalities of A-node and I-node. SI-nodes combine the functionalities of S-node and I-node. SA-nodes, such as Internet appliances, combine the functionalities of S-node and A-node. The CP provides service management for the A-node, S-node and I-node. The CP has the following functionalities:

- *A-node management:* It supports A-nodes with power saving and provides data access between the A-node and the Internet or PCs on other network segments.
- *S-node management:* It provides session setup for the S-node and assigns priority channel access to the S-node.
- *I-node management:* It provides connection setup for the I-node, allocates the I-node dedicated network resources (i.e., bandwidth), and provides connectivity between the I-node and the Public Switch Telephone Network (PSTN).

The HomeRF standard defines two main network topologies: Ad hoc Network and Managed Network. An ad-hoc network (see Figure 6.1) is a distributed network (similar to the ad hoc network concept in IEEE 802.11) in which HomeRF devices can establish peer-to-peer communication. An ad hoc network includes only A-nodes. There is no CP in an ad hoc network. A-nodes in an ad hoc network cannot operate in the power saving mode.

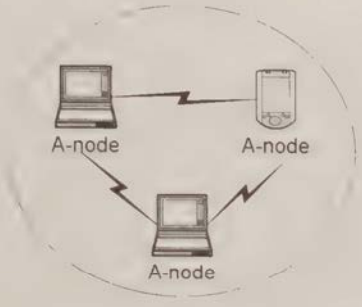


Figure 6.1 Ad Hoc Network

The Managed Network (see Figure 6.2) is a network managed by the CP. The HomeRF devices can either establish peer-to-peer communication with each other or establish communication through the CP depending on the type of the HomeRF devices, for example:

- *A-node:* Peer-to-peer communication between the A-nodes, or communication to the Internet or PCs on other network segments through the CP.

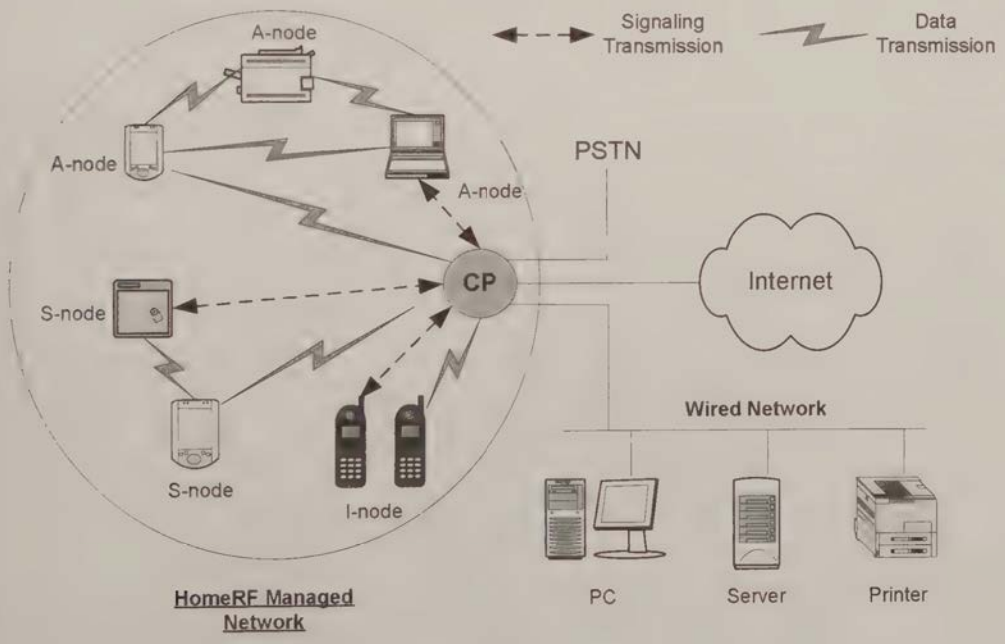


Figure 6.2 Managed Network

- *S-node:* Peer-to-peer data communication between the S-nodes, or communication with the CP for signaling purposes (i.e., session setup, priority assignment) or for data transfer to the Internet (i.e., streaming video from the Internet).
- *I-node:* Communication to PSTN or PC through the CP for transmission of both data and signaling. Peer-to-peer communication between I-nodes is not allowed.

6.2.2 Protocol Stack

As mentioned earlier, HomeRF protocols use best effort asynchronous data service, priority asynchronous data service, and isochronous data service to support simultaneous sessions for audio, video, and data. Both the asynchronous and priority asynchronous data services are packet switched using a carrier sense multiple access (CSMA/CA) media access control (MAC). The isochronous data service is for voice that requires strict jitter and latency and is supported by a circuit switched, time division multiple access (TDMA) MAC. HomeRF supports TCP/IP and User Datagram Protocol (UDP/IP) traffic as well as Digital Enhanced Cordless Telecommunications (DECT) traffic (see Figure 6.3). DECT, which is a cordless phone standard mainly used in Europe, uses a TDMA based MAC (i.e., assigns users with predetermined time slots). DECT provides speeds of up 2 Mbps. The implementation of the HomeRF MAC differs from the DECT MAC. However, because the services it provides are similar, the DECT higher layers can be hosted on the HomeRF MAC with minimal modification. The use of the DECT protocol enables HomeRF to support call setup for the isochronous data service and to interoperate with the PSTN.

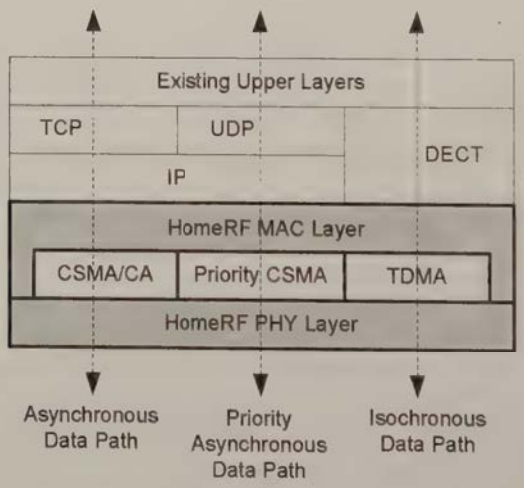


Figure 6.3 HomeRF Protocol Stack

6.3 Physical Layer

The PHY layer is based on Frequency Hopping Spread Spectrum (FHSS) and operates in the 2.4 GHz frequency band. It provides 75 1 MHz channels for 1.6 Mbps data and voice communication, and 15 5 MHz channels for 5 and 10 Mbps data communication.

HomeRF specifies an adaptation mechanism to ensure that two adjacent hops will not both be within the frequency range where interference may occur. When interference is detected, the hopping series set is examined to find out if two consecutive hops are both within the interference range. If such hops are found, then an attempt is made to switch these hops with hops that are outside the interference range. For example, if a microwave oven interferes, the hopping series will be adjusted to avoid the microwave frequency.

6.4 Media Access Control (MAC)

The standard organizes the communication channel into a superframe structure that includes two periods—contention period and contention-free period. The duration of the superframe is fixed and has two possible values: 20 ms when the superframe contains only a contention period (see Figure 6.4A) and 10 ms when the superframe contains both contention and contention-free periods (see Figure 6.4B). Each superframe begins with a hopping period (HOP) where HomeRF devices change their hopping frequency. When there is a contention-free period, the superframe contains two subframes—contention-free period for voice transmissions and optional contention-free period for voice retransmissions (see Figure 6.4B).

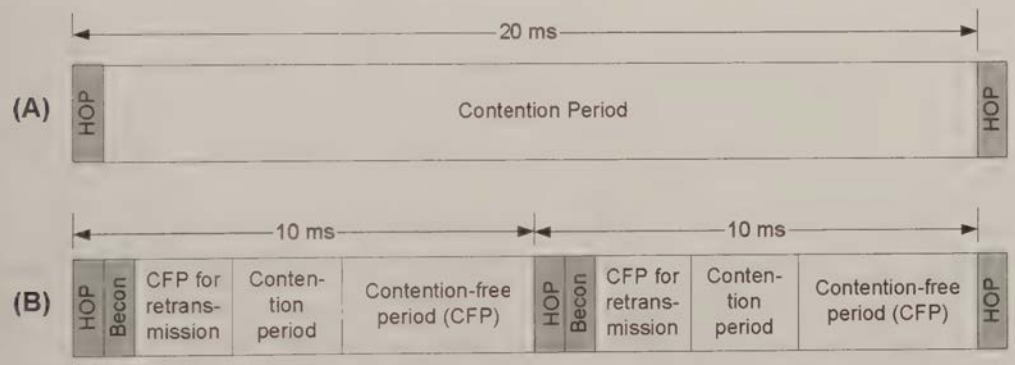


Figure 6.4 HomeRF Superframe Structure

6.4.1 Channel Access

The standard defines two channel access schemes for each period (contention and contention-free):

1. *Contention period:* The channel access is CSMA/CA with added priority access and time reservation mechanisms. The priority access feature provides priority access to the streaming media transmission over the asynchronous data transmission. Therefore, this channel access is used by A-nodes and S-nodes.

2. *Contention-free period:* The channel access is TDMA with the re-transmission option. Since this channel access provides isochronous data transmission service (i.e., voice), it is used by the I-nodes (e.g., cordless phones).

As shown in Figure 6.5, the contention period includes asynchronous data transmission where the streaming media sessions receive priority. The contention-free period includes voice transmissions and optional voice retransmissions.

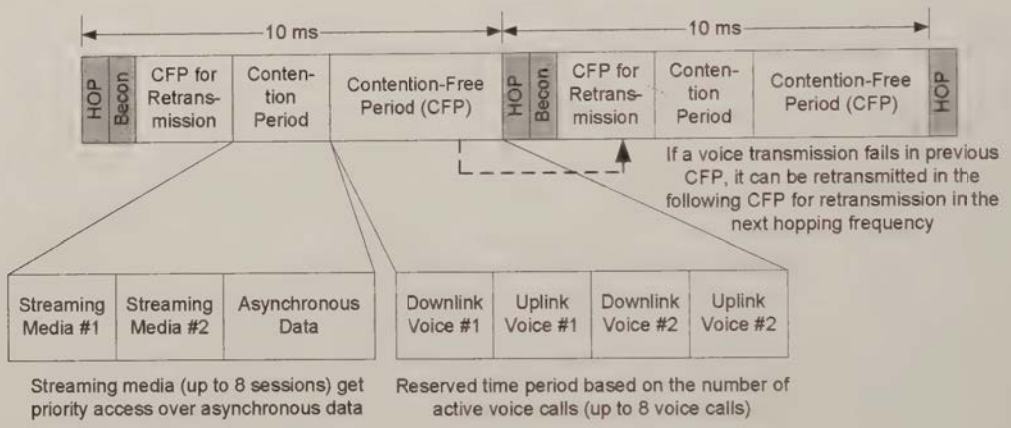


Figure 6.5 Channel Access

6.4.2 Contention-based CSMA/CA with Priority Access and Time Reservation

Contention-based CSMA/CA with priority access and time reservation supports two modes of asynchronous data access: legacy non-prioritized data access and priority asynchronous data access. In case of non-prioritized access, stations that have data to transmit access the wireless medium in an uncoordinated way. Consequently, this access has the lowest access priority. In the priority access, the stations access the wireless channel based on a priority access value determined by the CP. S-nodes will be granted priority access whereas A-nodes will access through the non-prioritized service. The CSMA/CA mechanism supports a multi-rate data service of 800 kbps, 1.6 Mbps, 5 Mbps, and 10 Mbps.

Similar to IEEE 802.11 (see Chapter 4), stations avoid collisions by listening to transmissions on the wireless medium. The station will transmit only if it senses that the medium is free for a certain period of time referred to as Interframe Space (IFS). Then, the station randomly selects a time slot within the contention window to start transmission. A collision is discovered by the fact that the sending station does not receive the acknowledgment from the receiving station. In this case, the station will double the contention window size and retransmit data by reselecting the time slot from the new contention window. The IFS is similar to IEEE 802.11:

- SIFS (short IFS) is the shortest time between two consecutive packet transmissions and used when there is no need to perform a Clear Channel Assessment (CCA) procedure before initiating a transmission.
- DIFS is the shortest time between the last bit of a sequence of a complete data transmissions and the first bit of a new packet sequence. This interval includes the time to recover from any prior transmission, perform a CCA procedure, and start a transmission. DIFS definition in HomeRF standard is slightly different from DIFS definition in IEEE 802.11. DIFS in HomeRF includes the randomized time slot duration of the contention window, which is not included in IEEE 802.11.

In addition, this MAC supports time reservation which enables the uninterrupted transmission of a sequence of multiple data packets. Stations that listen to the exchange of time reservation transmissions will not attempt transmissions. Hence, providing a contention-free medium is for the stations participating in this time reservation exchange. Time Reservations are used to signal that the time reservation is made, canceled, or requested. An optional Request-To-Send (RTS) and Clear-To-Send (CTS) packet exchange (similar to IEEE 802.11) can be used for the reservation process in case of hidden nodes. In the hidden node case, the RTS-CTS exchange ensures that packets at the source vicinity as well as the destination vicinity will defer for transmission during the reserved period of time.

The standard also defines CSMA/CA with priority access (see Figure 6.6), which can support up to eight priorities. Priority access can be achieved through reserved time slots (slot #1 up to slot #8) in the contention window. The CP assigns these slots. Priority access is used by streaming media applications. A streaming media application with stricter QoS requirements will be assigned a lower slot number which translates in a shorter DIFS. Other asynchronous data sessions select the slot number of the contention window excluding the reserved slots. Therefore, the streaming media with the lowest assigned slot number will have the highest priority access. Because the channel access scheme uses the CSMA/CA mechanism, the time delay is variable and cannot be guaranteed.

6.4.3 Reserved Time TDMA with Retransmission Option

Reserved time TDMA with re-transmission option operates in the contention-free period and is available only in a CP managed wireless network. The CP manages the I-nodes' access to the wireless medium by sending special control packets, or beacons, that contain TDMA information. Because all I-nodes' data transmission gets through the CP, there are two directions of transmission: uplink transmission (from I-node to CP) and downlink transmission (from CP to I-node). This MAC mode is connection oriented—that is, each voice call originated at an I-node is required to establish a connection with the CP. The CP will allocate a downlink-uplink pair of time slots to each call.

Figure 6.7 shows a potential scenario of HomeRF operation. Figure 6.7A shows the channel allocation when only data are communicated using the contention based protocol. The cycle is 20 ms long and is used by several data packets and their respective acknowledgments (ACKs)

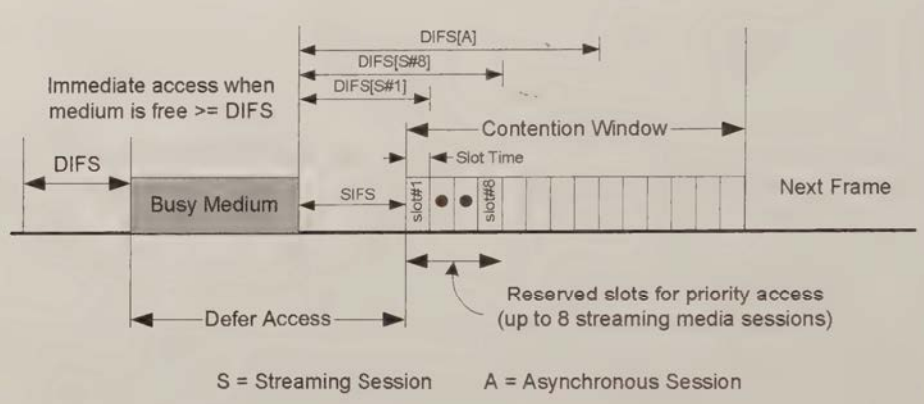


Figure 6.6 CSMA/CA with Priority Access

(Figure 6.7A). In Figure 6.7B, two voice conversations are added and the cycle is reduced to 10 ms. The Beacon, which is a management control packet, starts each cycle and allocates downlink and uplink TDMA time slots. After each cycle the Frequency Hopping radio hops to the next frequency of the frequency series.

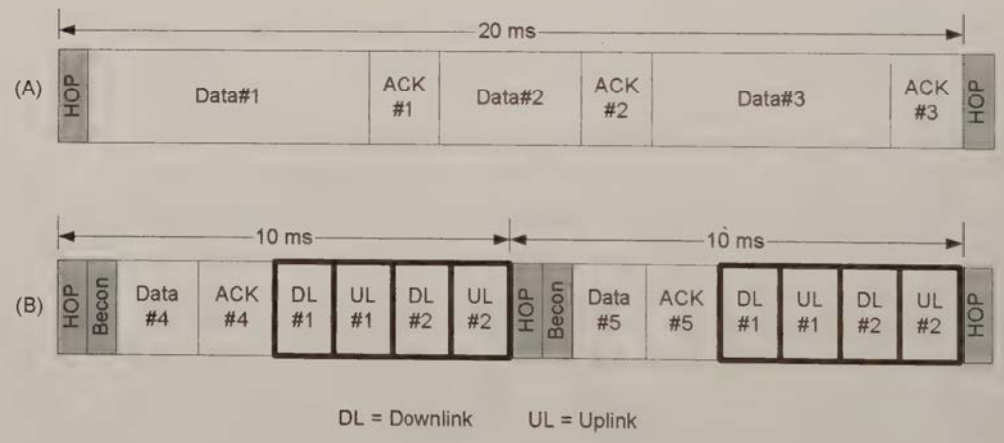


Figure 6.7 Example of HomeRF Operation

The connection-oriented TDMA protocol is referred to as the "single retransmission after hop" retry protocol (Figure 6.8). This protocol supports high quality voice transmissions (e.g., DECT based devices) that request bounded delays. The CP allocates time slots to every active connection. This allocation may change if other connections are terminated. The TDMA acknowledgment is a simple piggyback acknowledgment inserted in a later TDMA packet in the reverse direction. If the CP does not receive a valid TDMA packet or if the TDMA packet does

not include an acknowledgment, then the CP schedules a retry time slot for the connection. This protocol provides bounded time delay because each time sensitive connection is provided dedicated bandwidth in both directions (from/to the CP) (i.e., dedicated TDMA slots in every cycle) and dedicated bandwidth in the succeeding frame if a retransmission is required. A retransmission, if needed, is done in the consecutive cycle which is less than 10 ms later.

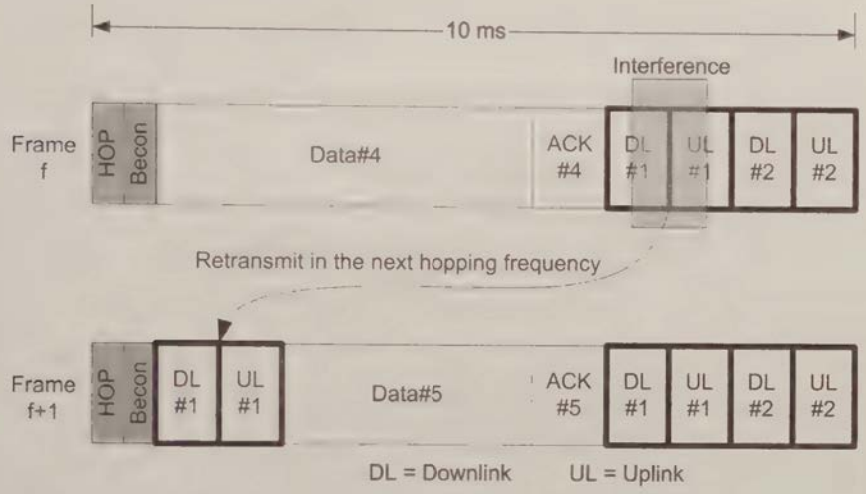


Figure 6.8 TDMA with Retransmission

This protocol also provides support for a connectionless broadcast service used by broadcasting applications whose packets do not require acknowledgments. Such applications include I-node pages, cadence ringing, caller ID, and voice announcement. This service is provided by only using the downlink time slots.

6.5 QoS Support

HomeRF standard explicitly defines the QoS service it can support. As mentioned at the beginning of this chapter, HomeRF offers three main services: asynchronous data service, priority asynchronous data service, and isochronous data service. In this section, we elaborate the QoS aspects of each service. We will describe the service characteristics and the QoS mechanisms involved.

6.5.1 Isochronous Data Service

This service is designed for applications (i.e., PSTN call, Intercom call, PC call, Conferencing) that originate at an I-node which require strict QoS control in terms of bandwidth, delay, and delay jitter. To achieve this service, the following service characteristics and QoS mechanisms are involved:

1. *Connection-oriented service:* An I-node sends the connection request to the CP. The CP assigns the connection ID and allocates the slot position in the contention-free period. This information is included in the Beacon. The CP identifies the connection by the connection ID (per-flow classification).
2. *Delay and delay jitter:* Isochronous data service provides bounded delay and delay jitter. Bounded delay and delay jitter are achieved using a TDMA based channel access scheme with retransmission option. The CP allocates a fixed transmission duration for each connection.
3. *Reliability:* The service allows the retransmission option to increase the transmission reliability.
4. *Priority:* This service has the highest priority.

In conclusion, the isochronous data service can provide quantitative QoS services.

6.5.2 Priority Asynchronous Data Service

This service, which is designed for applications (i.e., streaming video and audio) that originate at S-nodes, involves the following service characteristics and QoS mechanisms:

1. *Session-oriented service:* An S-node sends the connection request to the CP. The connection request contains a priority value (0-7 value which is similar to the IEEE 802.1D priority definition) or a delay/delay jitter requirement. The CP assigns the session ID (SID) and determines the associated priority based on the priority value or delay/delay jitter requirement information. The classification is based on the SID.
2. *Delay and delay jitter:* Delay and delay jitter are not bounded but are less than the values provided by the asynchronous data service. Priority CSMA/CA channel access is used. Packet scheduling follows a strict priority policy (i.e., a session with a higher priority will receive service first).
3. *Priority:* This service has higher priority than asynchronous data service but lower than isochronous data service.

From the service characteristics presented above, it can be concluded that priority asynchronous data service delivers qualitative (relative) QoS service. The standard does not define traffic policing mechanism. A higher priority session can occupy and starve the bandwidth of lower priority sessions and asynchronous data service.

6.5.3 Asynchronous Data Service

This service is designed for applications (i.e., non-real-time applications) that originate at A-nodes and do not have any specific QoS requirements. This service has the following service characteristics and QoS mechanisms:

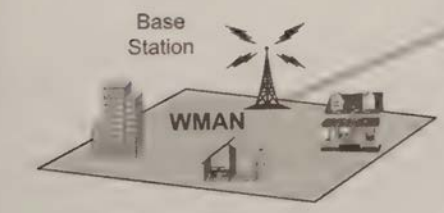
1. *Connectionless service:* This service does not require connection setup.
2. *Delay and delay jitter:* Since the channel access mechanism uses CSMA/CA, the delay and delay jitter are variable.
3. *Priority:* This service has the lowest priority.

In summary, asynchronous data service delivers best effort services.



PART 3

Wireless Metropolitan Area Networks





CHAPTER 7

IEEE 802.16

7.1 Introduction

IEEE has been working on the standard for MAC and physical layer specifications for wireless metropolitan area networks (WMANs or WirelessMAN™). The goal of WMANs is to provide high-speed wireless Internet access similar to wired access technologies such as cable modem, digital subscriber line (DSL), Ethernet, and fiber optic. IEEE was motivated by the ability of the wireless technologies to cover large geographic areas without the need to deploy wires.

IEEE published the first standard in 2002. The first approved standard, referred to as IEEE 802.16.1, is geared towards network access between buildings with exterior antennas communicating with a central radio base station (BS) in the 10 to 66 GHz frequency bands. In parallel, the European Telecommunications Standards Institute (ETSI) has been working on a standard that serves similar goals. This effort, referred to as HiperAccess, was completed in 2002, when ETSI decided to adapt IEEE 802.16.1 as its HiperAccess standard.

Since several WMANs can coexist and consequently interfere with each other, the IEEE organization also published in 2001 a recommended practice, referred to as IEEE 802.16.2 for Local and Metropolitan Area Networks, such that mutual interferences between coexisting WMANs are minimized.

IEEE is still working on an addition to the 802.16.1 standard that details MAC and physical layer specifications for the 2 to 11 GHz bands. This effort is referred to as IEEE 802.16a. The European organization, ETSI, has been working on a similar project, referred to as HiperMAN. The organizations are working closely together and it is expected that IEEE 802.16a will be adapted as HiperMAN.

IEEE 802.16 is also working to introduce an amendment, referred to as P802.16c, that will detail System Profiles for the 10 to 66 GHz band. These system profiles help in the deployment of interoperable systems.

The IEEE 802.16 group realized the need for multimedia applications and the required QoS support. Therefore, IEEE 802.16 has included a number of QoS signaling mechanisms. However, the algorithms that use such signaling mechanisms in order to provide QoS support are vendor specific and are left out of the standard. This allows vendors to differentiate their products but still be interoperable.

7.2 IEEE 802.16.1

IEEE 802.16 was designed considering that a metropolitan area can include hundreds or thousands of subscribers which require high-speed connections. The MAC is built to accommodate a point to multipoint topology, reflecting the fact that the BS serves multiple subscriber stations (SSs). The 802.16a MAC will add the capability of serving mesh networks. The MAC addresses the high-speed QoS requirements with a flexible design of uplink (SS to BS) and downlink (BS to SS) channels. The BS has full control on the bandwidth allocation on both channels. Access allocation is provided by the BS via a request-grant mechanism, in which the SS requests access and the BS grants access. The standard realizes that users may have a variety of needs resulting from their legacy voice systems, voice over IP systems, and their TCP/IP packetized systems. It also realizes that, due to different applications, users may execute applications that result in either continuous or bursty traffic, with different QoS requirements.

7.2.1 Architecture

7.2.1.1 Network Topology

The IEEE 802.16.1 is geared toward point to multipoint (PMP) communication topology, in which a BS communicates with numerous stations, or SSs. All connections need to go through the BS (i.e., there is no direct communication between the SSs). Due to the short wavelength of the 10 to 66 GHz frequency bands, with channels of 25 or 28 MHz, the standard requires line of sight between the BS and SS. Data rates in excess of 120 Mbps can be supported. The IEEE 802.16a extension effort supports significantly reduced data rates at 2 to 11 GHz frequency bands. This approach does not require line of sight and supports mesh topology in which connections do not need to go through the BS.

Figure 7.1 shows an example of a network topology in which PMP and mesh topologies are used to cover a large metropolitan area. The in-building communication can be provided through wireless LANs (e.g., IEEE 802.11) or via an implementation of the 802.16a extension. Hence, IEEE 802.16 can provide a complete wireless communication solution for the targeted metropolitan area.

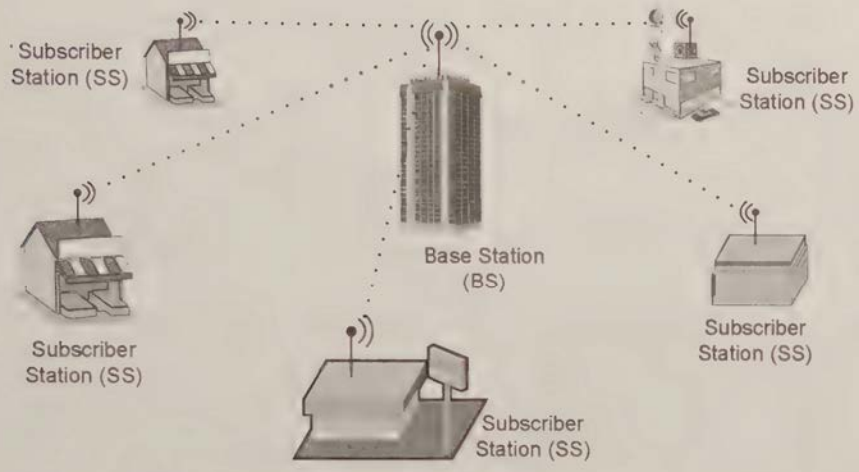


Figure 7.1 Point to Multipoint Network Topology

7.2.1.2 Protocol Stack

The standard is divided into various layers as shown in Figure 7.2. The MAC layer is divided into three sublayers: The Service Specific Convergence Sublayer (CS), the MAC Common Part Sublayer (CPS), and the Security Sublayer.

The Service Specific Convergence Sublayer (CS) transforms incoming network data, received through the CS service access point (SAP) into MAC data packets, preserving or enabling QoS, and allowing bandwidth allocation. This transformation maps external network information into IEEE 802.16 MAC information, such as service flow and connection ID. The current standard details two Convergence Sublayer specifications for ATM CS and for Packet CS. The Packet CS is defined for mapping external packets such as IPv4, IPv6, Ethernet, and virtual local area network (VLAN).

The MAC Common Part Sublayer (CPS) provides access control functionality, bandwidth allocation, connection establishment, and connection maintenance.

The Security Sublayer provides security services including authentication, key exchange, and encryption. Data, physical layer (PHY) control, and management information are exchanged between the MAC CPS and the PHY via the PHY SAP. The PHY layer includes multiple radio options.

7.3 Physical Layer

The IEEE 802.16 defines the physical layer for the 10 to 66 GHz frequency bands assuming line of sight between the BS and SS. It uses burst single-carrier modulation in which all data are sequentially transmitted in a single frequency. The standard allows the use of directional antennas to increase capacity and reduce interference from adjacent transmissions. The physical layer supports wide channel bandwidth of 20 or 25 MHz in the U.S. and 28 MHz in Europe. With 20

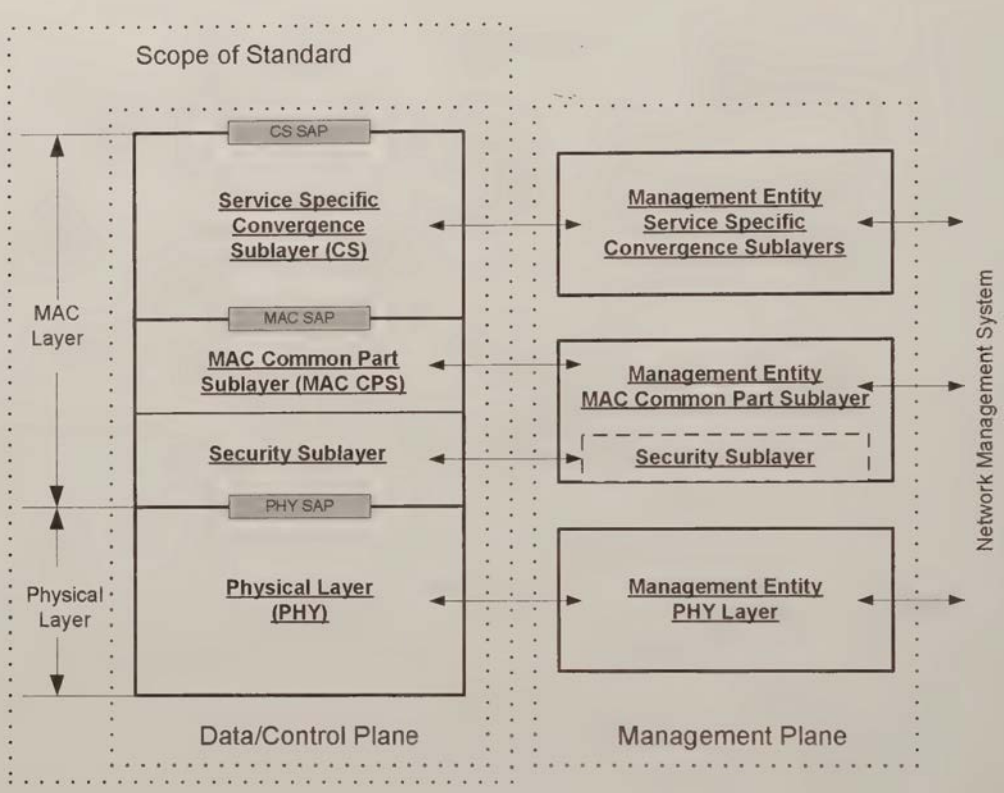


Figure 7.2 IEEE 802.16 Protocol Layer

MHz channel width vendors can support up to 96 Mbps, with 25 MHz up to 120 Mbps, and with 28 MHz up to 134 Mbps. The communication path between an SS and the BS has two directions: uplink, from the SS to the BS, and downlink, from the BS that may reach many SSs. Both uplink and downlink can operate in different frequencies using Frequency Division Duplexing (FDD) or share the same frequencies using Time Division Duplexing (TDD). In both FDD and TDD systems, the uplink and downlink channels are structured into frames. In TDD, the frame is divided into two subframes, uplink and downlink subframes, where the uplink subframe follows the downlink subframe (Figure 7.3). In FDD, the uplink and downlink subframes are concurrent in time but are carried on separate frequencies (Figure 7.4). Each subframe (in TDD or FDD) consists of a number of time slots in which both TDD and FDD systems deploy TDM transmission. All SSs and the BS have to be synchronized and transmit data bursts into predetermined time slots.

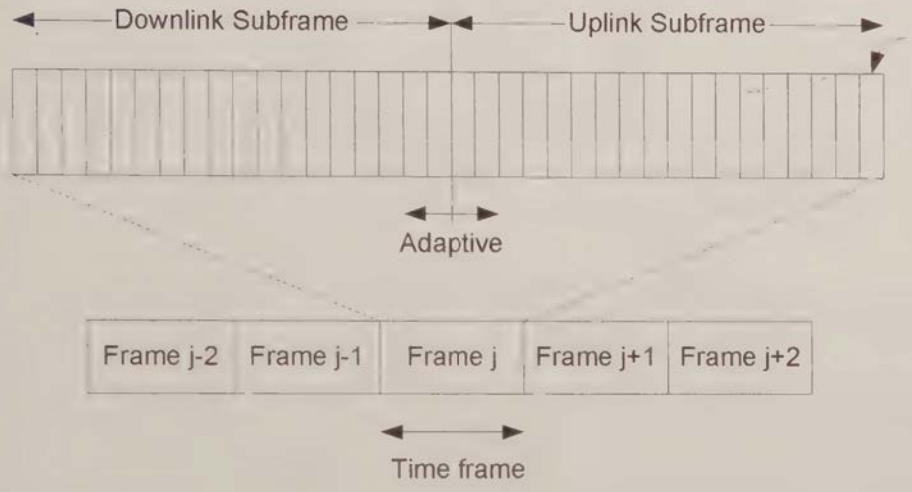


Figure 7.3 TDD Frame Structure

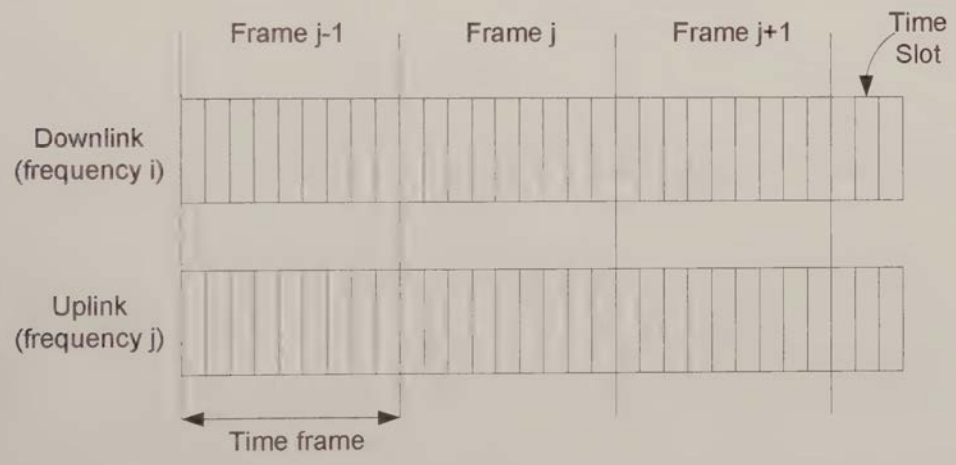


Figure 7.4 FDD Frame Structure

The standard also supports full duplex subscriber stations (which can transmit and receive simultaneously) and half duplex subscriber stations (which can either transmit or receive). For FDD systems, the time slot assignment is different for full or half duplex subscriber stations. When there are half duplex subscriber stations, the BS cannot assign an uplink period for an SS which overlaps with the SS downiink period. For TDD systems, due to the fact that there is no overlapping period between the uplink and downlink channels, both full and half duplex subscriber stations can have the same time slots assignment.

7.3.1 Adaptive Data Burst Profiles

The standard supports adaptive data burst profiling in which transmission parameters, such as modulation and Forward Error Correction (FEC) coding settings, can be modified individually to each SS on a frame-by-frame basis in both uplink and downlink transmissions. The standard allows three types of modulation schemes: Quadrature Phase Shift Keying (QPSK), 16 Quadrature Amplitude Modulation (QAM), and 64 QAM. Different combinations of modulation and FEC provide different transmission robustness and transmission speed (i.e., 64 QAM delivers high transmission speed but is prone to interference; whereas, QPSK delivers low transmission speed but is tolerant to interference). Data burst profiles, which include parameters such as radio modulation type and FEC, are identified by a code called Interval Usage Code (IUC). There are two types of IUC: 1) Downlink Interval Usage Code (DIUC), which identifies the downlink data burst profiles, and 2) Uplink Interval Usage Code (UIUC), which identifies the uplink data burst profiles. DIUC is included in a MAC message called Downlink Channel Descriptor (DCD) message; whereas, UIUC is included in the Uplink Channel Descriptor (UCD) message. DCD and UCD messages are transmitted periodically by the base station in order to define the downlink and uplink channel characteristics, respectively.

7.4 Media Access Control (MAC)

The 802.16.1 MAC is connection oriented. All traffic including inherently connectionless traffic is mapped into a connection. This provides the ability to request bandwidth with its associated QoS and traffic parameters for every connection. Connections are identified by Connection Identifiers (CIDs). The standard allows bandwidth provision to these connections to be either continuous or on demand. Each connection is associated with a service flow (SF) that is a unidirectional flow of MAC packets on a connection that is provided with a particular QoS. The service flow defines the QoS parameters for the packets that are transmitted on this connection. The Service Flow concept is central to the 802.16 MAC since it provides the mechanism for QoS and bandwidth allocation process. The MAC has also some reserved CIDs for purposes such as management, announcements, and broadcasting.

Established connections require active maintenance depending upon the type of service. For example, unchannelized T1 services require virtually no connection maintenance since they have a constant bandwidth allocation. Channelized T1 services require some maintenance due to the dynamic bandwidth requirements. Connections may be also terminated or adjusted based upon users' contracts. The standard supports all of these connection management functions through the use of static configuration and dynamic addition, modification, and deletion of connections.

7.4.1 Channel Access

As described earlier, the communication path has two directions—uplink and downlink. In uplink transmissions (SSs to BS), several SSs share the channel in a TDMA fashion. The Uplink Map Message (UL-MAP) is used to provide the channel access assignment to the subscriber

stations. The UL-MAP which is transmitted by the base station at the beginning of the frame defines the uplink channel access as well as the uplink data burst profiles (i.e., UIUC) in the current uplink subframe (see Figure 7.5B-C). The SSs are allowed to transmit data bursts at predetermined time slots as indicated in the UL-MAP. In downlink transmissions (BS to SSs), due to the fact that there is only one station (BS) transmitting the data, the channel access is rather simple. The data packets are transmitted by the BS to all SSs and an SS only picks up the packets destined to it. The Downlink Map Message (DL-MAP) is used to define the downlink data burst profiles (i.e., DIUC) in each time period in the current downlink subframe, in both TDD and FDD systems (see Figure 7.5A).

Both UL-MAP and DL-MAP indicate the starting time slot of each data burst (see Figure 7.6).

7.4.1.1 Downlink Subframe

The standard specifies different downlink frames for FDD and TDD systems.

The FDD downlink subframe (Figure 7.7) starts with a preamble, DL-MAP, and UL-MAP. The preamble helps in the physical layer transition and synchronization. The DL-MAP defines bursts' start times on the downlink for both TDM and time division multiple access (TDMA) methods. The downlink subframe is divided into TDM and TDMA portions, where the TDMA portions follow the TDM portions. The TDM portions are not separated by gaps or preambles since there is only one transmitter, the BS. The TDM portions contain data transmitted to one or more of the following: full duplex SSs, half duplex SSs scheduled to transmit later in the frame, and half duplex SSs not scheduled to transmit in this frame. On the other hand, the TDMA portions are separated by preambles and gaps. The TDMA portion is used to transmit data to any half duplex SSs scheduled to transmit earlier in the frame than they receive data. This allows an individual SS to decode a specific portion of the downlink without the need to decode the entire downlink subframe.

The TDD downlink subframe (Figure 7.8) starts with a preamble, DL-MAP, and UL-MAP. The preamble helps in the physical layer transition and synchronization. The TDD downlink subframe contains data packets that are transmitted to SSs and ends with a Transmit/Receive (Tx/Rx) time gap. In the TDD downlink subframe there are only TDM portions.

7.4.1.2 Uplink Subframe

The uplink subframe (Figure 7.9) is used by the SSs to transmit to the BS.

The uplink subframe includes the following three periods: Initial Maintenance period, Request Contention Opportunities period, and Scheduled Data grants period. The different periods are identified by their respective UIUC that is specific to an uplink period. The BS announces these periods and associated burst classes in the preceding downlink subframe's UL-MAP. The BS can specify such periods in any order and length. The BS can group the periods allocated to the Initial Maintenance and Request Contention Opportunities and leave the remaining periods for data transmission. We provide next a more detailed description of each period.

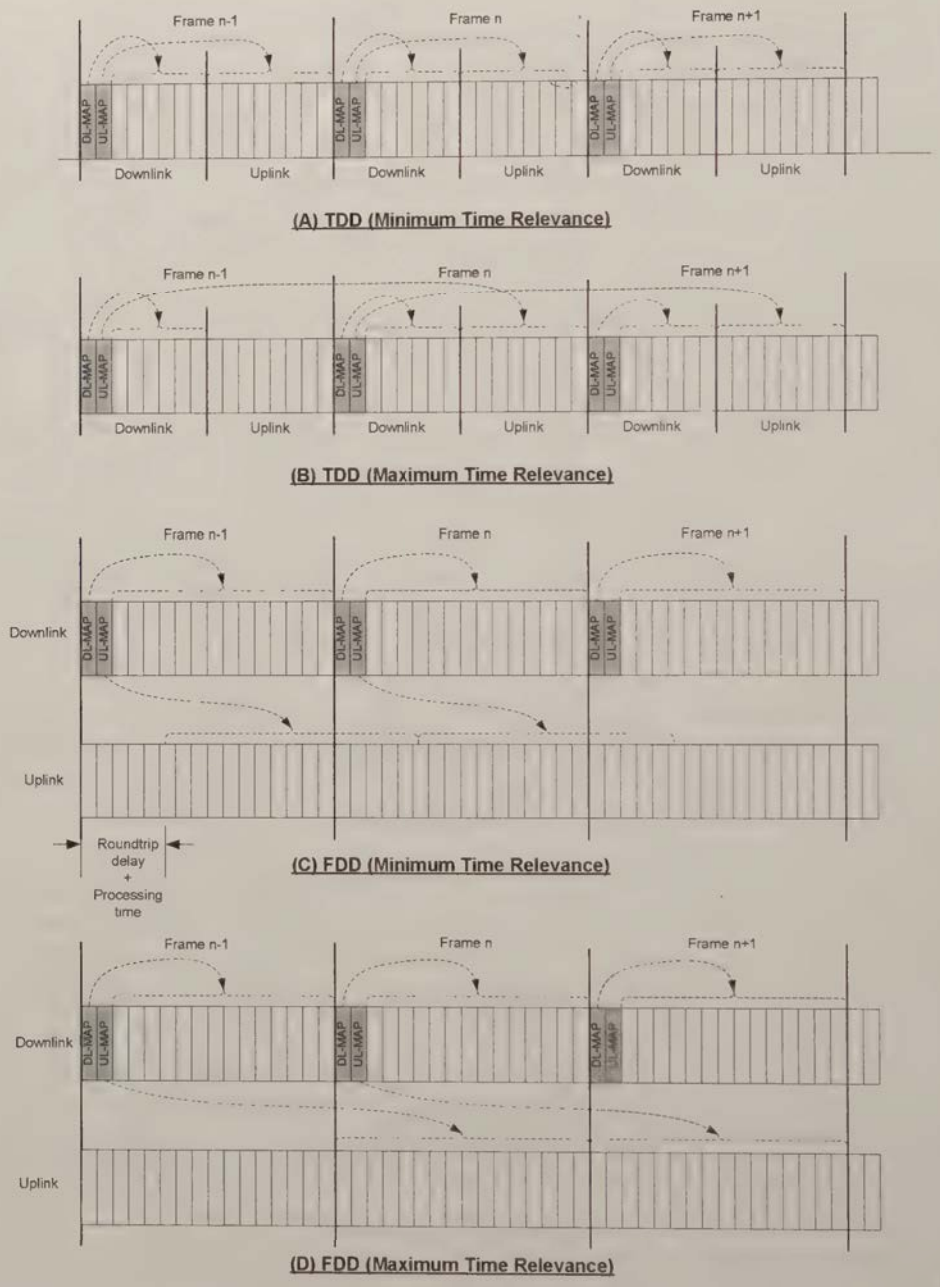


Figure 7.5 MAC Control Information

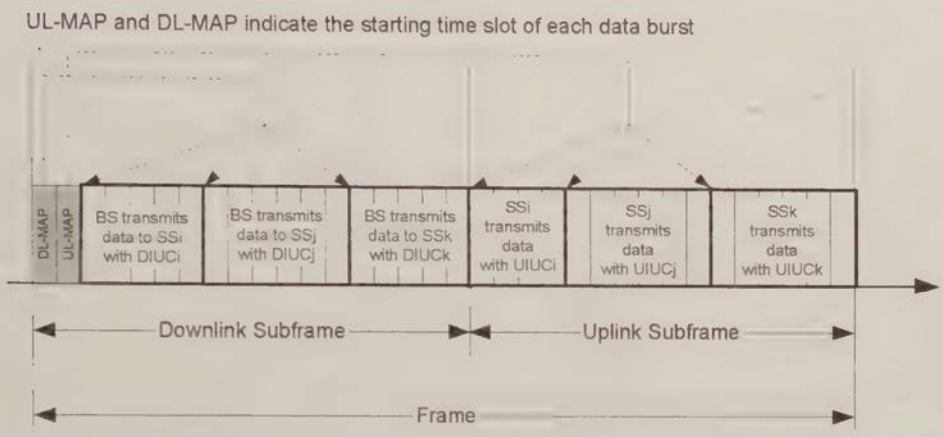


Figure 7.6 DL-MAP and UL-MAP

In the Initial Maintenance period, identified by UIUC = 2, the SS sends transmissions or bursts needed to carry out Initial Maintenance functions. For example, ranging requests are sent by the SS at the initialization phase and periodically thereafter. The BS uses such requests to determine network delay and request power or downlink burst profile changes. In this period also new stations may join the network. Since several SSs can access the channel simultaneously, collisions can occur in this period.

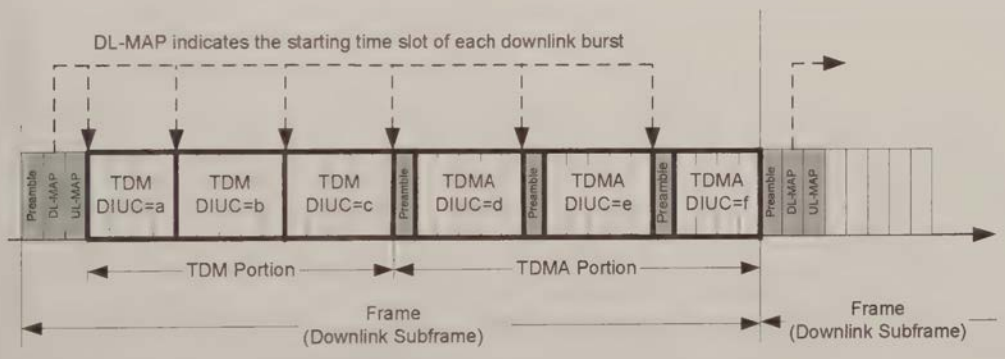


Figure 7.7 FDD Downlink Subframe Structure

In the Request Contention Opportunities period, identified by UIUC = 1, the SSs request bandwidth based upon multicast and broadcast polls by the BS. Since several SSs can access the channel simultaneously, collisions can occur in this period.

In the Scheduled Data grants period, identified by UIUC that defers from 1 or 2, the SS transmits data based on grants allocated by the BS. An SS transition gap separates each Scheduled Data period to allow transition from one period with a specific burst profile to another period which may have a different burst profile. Each such period starts with a preamble to allow the new SS to synchronize.

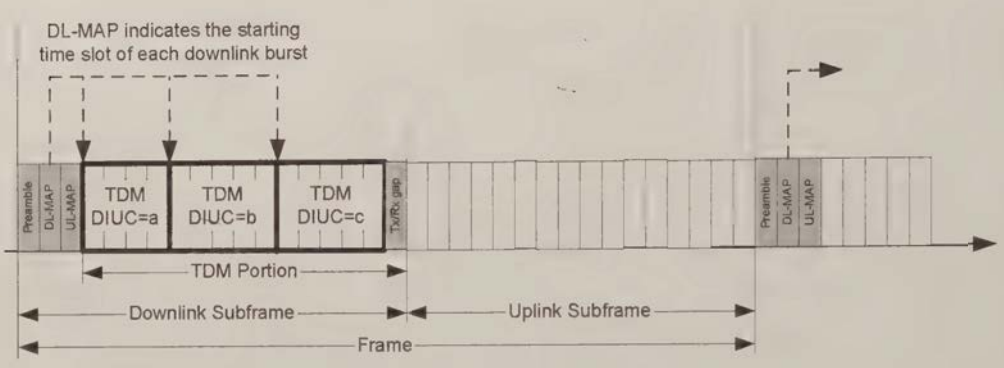


Figure 7.8 TDD Downlink Subframe Structure

7.4.1.3 Contention Resolution

Since several SSs can access the channel simultaneously during the Initial Maintenance and Request periods, their respective transmissions can collide. If such a collision occurs the transmitting SSs will go through a contention resolution process. It is possible that the SS will have multiple uplink service flows, each associated with its CID. The SS will consider these CIDs or their QoS needs in this process. The mandatory process by the standard is based on a truncated binary exponential backoff, with an initial backoff window and a maximum backoff window controlled by the BS. The algorithm is defined in the standard.

7.4.2 Bandwidth Request Mechanisms and Bandwidth Allocation

The standard defines various mechanisms for the SS to access the shared uplink and request transmission opportunities (bandwidth) and for the BS to grant such transmission opportunities. The key mechanisms of the request-grant process are the bandwidth request mechanism and the bandwidth allocation. These mechanisms enable multiple types of service flows which support a wide range of applications. First, we introduce the bandwidth request and the bandwidth allocation mechanisms and then we describe each type of service flow offered by the standard.

7.4.2.1 Bandwidth Request Mechanism

Bandwidth is always requested on a connection (i.e., CID) basis. Each connection in an SS requests bandwidth through an MAC message called BW Request message. BW Request can be sent in a stand-alone packet or piggyback on another packet. The request is based on the number of bytes needed to carry the MAC packet without considering the physical layer overhead. The requested bandwidth can be either incremental, meaning how much additional bandwidth is needed, or aggregate, meaning how much total bandwidth is needed. When the BS receives an incremental bandwidth Request, it adds the quantity of bandwidth requested to its current allocated bandwidth. When the BS receives an aggregate bandwidth Request, it replaces the allo-

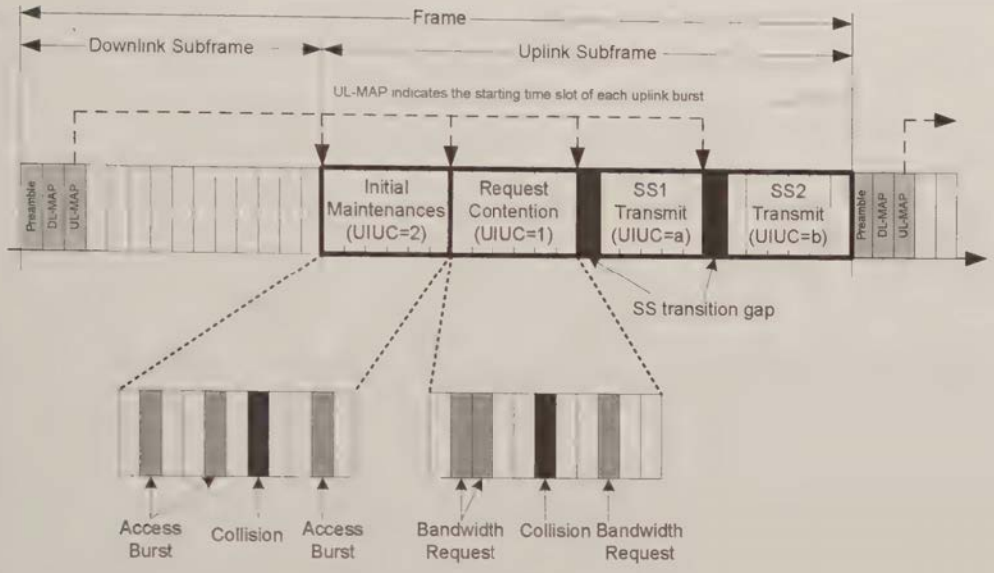


Figure 7.9 Uplink Subframe Structure

cated bandwidth with the quantity of bandwidth requested. With piggyback BW Request, the connection can only request incremental bandwidth. With stand-alone BW Request, the connection can request either incremental or aggregate bandwidth. BW Request can be initiated directly by the connection or initiated in response to receiving a Polling Message from BS. BW Request can be transmitted during the uplink subframe in two different ways: 1) Contention request opportunities in which BW Request is transmitted in Request Contention period and 2) Contention-Free Request Opportunities in which BW Request is transmitted in the predefined time slot of uplink subframe indicated by Information Element (IE) of UL-MAP or is piggybacked with other packets.

The polling mechanism is geared for detailed and flexible requests but requires more overhead. BS issues unicast polling packets for the SS as well as multicast and broadcast polls. The polling process does not include an explicit packet from the BS to the SS. Rather, the SS is allocated bandwidth for its potential request in the downlink subframe field that details the coming uplink transmission opportunities, referred to as uplink map.

The standard also allows the BS to poll stations in groups or as a whole. This mode can be used in cases where the BS decides that there is not enough bandwidth to support individual unicast polls or in cases where there are many inactive SSs. Certain connection CIDs are reserved for such multicast and broadcast polls. The SSs addressed in this group polling will need to contend during the uplink period to send their transmission requests.

7.4.2.2 Bandwidth Allocation

As mentioned in the previous section, each SS requests bandwidth on a connection basis. The BS grants or allocates bandwidth by deploying one of the following two modes: Grant Per Subscriber Station (GPSS) and Grant Per Connection (GPC). In GPC mode, the BS allocates bandwidth to individual connections (CID). In GPSS mode, the BS allocates bandwidth for individual SSs. In this case, the SS is responsible for distributing this allocated bandwidth to each connection within the SS. The SS decides which traffic should be transmitted first, considering each application's traffic needs and QoS requirements. Thus, GPSS mode requires more intelligence in the SS than does the GPC mode. This allows vendors to design different SS configurations. A vendor may choose a GPC design for simplicity. If the 10 to 66 GHz frequencies are used, the standard allows only SSs with GPSS mode.

The BS allocates bandwidth to a connection or a station by granting the transmission opportunity in the UL-MAP. The bandwidth allocation is mainly based on the following factors:

- The amount of bandwidth requested by the connections
- QoS parameters of delay and bandwidth needed by the current applications at the SS as reflected in the service flow parameters
- Available network resources

The BS can grant transmission opportunities for multiple connections to a single SS, each with a different QoS level. The BS may not grant bandwidth to the requesting SS due to reasons such as: the bandwidth requested by the SS is not available, the BS received a request that included errors, and the request did not go through due to collisions with other SSs or due to errors in the wireless media.

7.4.3 Service Flow Scheduling Service

The standard defines four types of service flows that provide QoS support for a wide range of applications. The services include: Unsolicited Grant Service (UGS), Real-Time Polling Service (rtPS), Non-Real-Time Polling Service (nrtPS), and Best Effort (BE) Service.

7.4.3.1 Unsolicited Grant Service (UGS)

The Unsolicited Grant Service (UGS) supports real-time service flows that generate fixed size periodic data packets, such as T1. This service provides periodic, fixed size grants that avoid the overhead and latency of frequent SS redundant requests. It also ensures that grants are available to meet the continuous needs of the service flow.

A connection with UGS service flow is prohibited from using any contention request opportunities. The BS will not provide any unicast request opportunities for this connection. The Information Element (IE) of the UGS service flow includes parameters such as the UGS size, nominal grant interval, the tolerated grant jitter, and the request/transmission policy. The SS updates the BS regarding the state of the UGS service flow, using special fields called Slip Indicator (SI) flag in the packet header. The SI flag indicates excessive queue length due to condi-

tions such as a lost allocation map or due to clock mismatch between the SS and the BS. In such cases the BS may grant additional bandwidth to compensate for bandwidth insufficiency. An SS that is granted UGS opportunity cannot "steal" bandwidth. In other words, the SS cannot use a portion of the allocated bandwidth to send another bandwidth request rather than sending data. Also it can not piggyback a request in the transmitted data. However, the SS can indicate that it needs a poll for non-UGS service by setting a special field, referred to as Poll-Me bit, in the packet header. This mechanism saves bandwidth, as the BS can poll such stations only if the Poll-Me bit is set. Once the BS receives this indication, it may follow with individual polling such as rtPS and nrtPS.

7.4.3.2 Real-Time Polling Service (rtPS)

The BS polls SSs with the purpose of allocating to these SSs bandwidth specifically for the purpose of making bandwidth requests. The rtPS supports real-time service flows that have periodic data packets of various sizes, such as MPEG streams. This service provides periodic, unicast request opportunities, to respond to the needs of the service flows. It allows the SS to dynamically specify the size of the requested grant. Compared to UGS, this service has additional overhead due to the polling process. However, it can handle variable grant sizes compared to UGS that can handle only fixed grant sizes. The standard specifies that the BS needs to poll the SS periodically in order to allow it to send unicast request opportunities. The polling frequency (the number of polls in a certain period of time) is not specified in the standard. In common practice, the BS should poll the SS frequently enough to meet the delay and bandwidth requirements of real-time applications. In this case, the SS is prohibited from using any contention request opportunities in order to avoid unpredictable delays. A connection with rtPS service flow is allowed to "steal" bandwidth if it is in the GPSS mode. rtPS service flow is also allowed to piggyback a BW Request on a data packet.

7.4.3.3 Non-Real-Time Polling Service (nrtPS)

The Non-Real-Time Polling Service (nrtPS) supports non-real-time service flows that have data packets of various sizes, such as FTP. The service provides consistent unicast request opportunities even during network congestion. The standard specifies that a BS needs to poll the SS on a regular basis (periodically or nonperiodically) and allow it to send unicast request opportunities. In this case, the SS is allowed to use contention request opportunities for the connection. In other words, the SS can simultaneously send unicast request packets in response to the BS Poll and content for request opportunities. A connection with nrtPS service flow is allowed to "steal" bandwidth if it is in GPSS mode. nrtPS service flow is allowed to piggyback a BW Request on a data packet.

7.4.3.4 Best Effort (BE) Service

The Best Effort (BE) Service supports service flows that do not require QoS support. The SS issues its requests in a contention period. An SS that is granted a request via BE is allowed to "steal" bandwidth if it is in GPSS mode. BE Service flow is allowed to piggyback a BW Request on a data packet.

The standard details extensive signaling techniques and access mechanisms for each service flow, but the details of bandwidth allocation, scheduling, and reservation management intelligence are out of the standard's scope and are left to be vendor specific. The ability to employ different combinations of these access mechanisms allows vendors to differentiate their products, tailor solutions to unique needs and users, optimize system performance, and use different bandwidth allocation algorithms while maintaining consistent interoperability. For example, contention may be activated to avoid polling of SSs that have been inactive for a long period of time.

7.4.4 Network Entry and Initialization

An SS wishing to enter the network has to go through the following steps:

1. Scan for a downlink channel and establish synchronization with the BS. The SS can either recall the downlink channel from its memory or continuously scan possible channels of the downlink frequency band until it finds a valid downlink signal. After it finds a downlink channel it needs to synchronize and attempt to acquire the channel control parameters for the downlink by searching for the DL-MAP sent by the BS.
2. Obtain transmit parameters. The SS searches for an uplink Channel Descriptor message from the BS in order to retrieve the transmission parameters for a possible uplink channel.
3. Adjust local parameters (e.g., transmit power) based on approved values or messages from the BS.
4. Negotiate basic capabilities. The SS informs the BS of its basic capabilities by sending an appropriate message to the BS. The BS acknowledges this message.
5. Authorize SS and perform key exchange. The BS goes through a process of authorizing the SS to enter the network and exchanges security keys with the SS.
6. Perform registration. The BS sends additional management messages and the SS becomes managed by the BS.
7. Establish IP connectivity. The SS receives an IP address from the BS.
8. Establish time of day. The SS and BS need to have equal timing information.
9. Transfer operational parameters. The BS sends additional configuration information to the SS.

7.5 QoS Support

The principal mechanism of IEEE 802.16 standard for providing QoS support is to associate a packet with a service flow. A service flow is a unidirectional flow of packets that provides a particular QoS. The standard details the mechanisms of how to allocate bandwidth and how to send the BW Requests in each service flow as described in the previous section. To summarize, the service flows and the supported applications are listed below:

- UGS Service flow supports real-time applications with constant bit rate (CBR) such as voice over IP and circuit emulation.
- rtPS Service flow supports real-time applications with variable bit rate (VBR) such as streaming video and audio.
- nrtPS Service flow supports non-real-time applications which require better service than BE, such as high-bandwidth FTP.
- BE Service flow supports applications that do not have any QoS requirements.

Each network application, first, has to register with the network. The network will associate the application with a service flow by assigning an unique Service Flow ID (SFID). All packets must be tagged with this assigned SFID in order for the network to provide the appropriate QoS. When the application wants to send data packets, it is required to establish a connection with the network and receives a unique CID assigned by the network. Therefore, the IEEE 802.16 data packets include both CID and SFID.

Next we describe QoS support from the following two aspects: QoS provision and QoS mechanisms.

7.5.1 QoS Provision

IEEE 802.16 broadband wireless access standard aims to provide fixed wireless access between the subscriber station (residential or business customers) and the Internet Service Provider (ISP) through the BS. The service model involves the following steps:

- In order to receive network service, the customers have to subscribe or register with the ISP or another entity that has the authority to control the network.
- Whenever the customers want to use the service (i.e., transmit data), they need to establish the connection with the network.
- If customers require QoS support for their applications, they are required to include a QoS parameter set that can be defined in any of the following ways: explicitly specifying all traffic parameters, indirectly referring to a set of traffic parameters via specifying a Service Class Name, or specifying a Service Class Name along with modifying the parameters.
- Considering this QoS parameter set and the available network resources, the network decides if it can provide the appropriate QoS support.

The standard's set of tools that support QoS for both uplink and downlink traffic include: configuration and registration functions for service flows, a signaling function for dynamically establishing QoS based service flows and traffic parameters, scheduling and QoS traffic parameters for uplink and downlink service flows, grouping of service flow properties into service classes to allow grouping of requests.

The standard's QoS provision, based on the "envelopes" model shown in Figure 7.10, defines several sets of QoS parameters:

- *ProvisionedQoSParamSet:* A set of external QoS parameters provided to the MAC, for example, by the network management system.
- *AdmittedQoSParamSet:* A set of QoS parameters for which the BS and possibly the SS are reserving resources since the associated service flows have been admitted by the BS. Resources are not limited to bandwidth and can include resources such as memory.
- *ActiveQoSParamSet:* A set of QoS parameters that reflect the actual service being provided to the associated active service flows.

The standard assumes an Authorization Module in the BS that approves or denies every change of QoS parameters associated with a service flow. This module can provision a service flow immediately or can defer the service flow activation to a later period.

The standard presents an "envelope" model that limits the possible values of the AdmittedQoSParamSet and ActiveQoSParamSet. The standard recognizes two models: Provisioned Authorization Model and Dynamic Authorization Model. In the Provisioned Authorization Model the parameters are provided, for example, by the network management system. In the Dynamic Authorization Model, the authorization module issues its decisions based on its vendor specific implementation. This implementation may use routines that communicate to external policy servers for additional information and input. The relationships among the QoS Parameter Sets in each model are shown in Figure 7.10. The ActiveQoSParamSet is always a subset of the AdmittedQoSParamSet, which is always a subset of the authorized "envelope." In the Dynamic Authorization Model, this envelope is determined by the Authorization Module (labeled as the AuthorizedQoSParamSet). In the Provisioned Authorization Model, this envelope is determined by the ProvisionedQoSParamSet.

7.5.2 QoS Mechanisms

In this section, we revisit the service flow and its mechanisms, emphasizing the QoS aspects.

7.5.2.1 Classification

All packets generated by active applications are tagged with CID and SFID. The classification module identifies the packets based on these tags and forwards them to the corresponding queues (see Figure 7.11).

Individual applications can establish individual connections. IEEE 802.16 can provide per-flow classification which supports per-flow QoS services.

7.5.2.2 Channel Access

As described in Section 7.5.1, the channel access method uses TDM for the downlink and TDMA for the uplink. The channel access is controlled by the BS through the UL-MAP and DL-MAP. Both TDM and TDMA are collision-free channel access schemes which provide tight channel access control. Therefore, these schemes can provide QoS support for applications with strict QoS requirements.

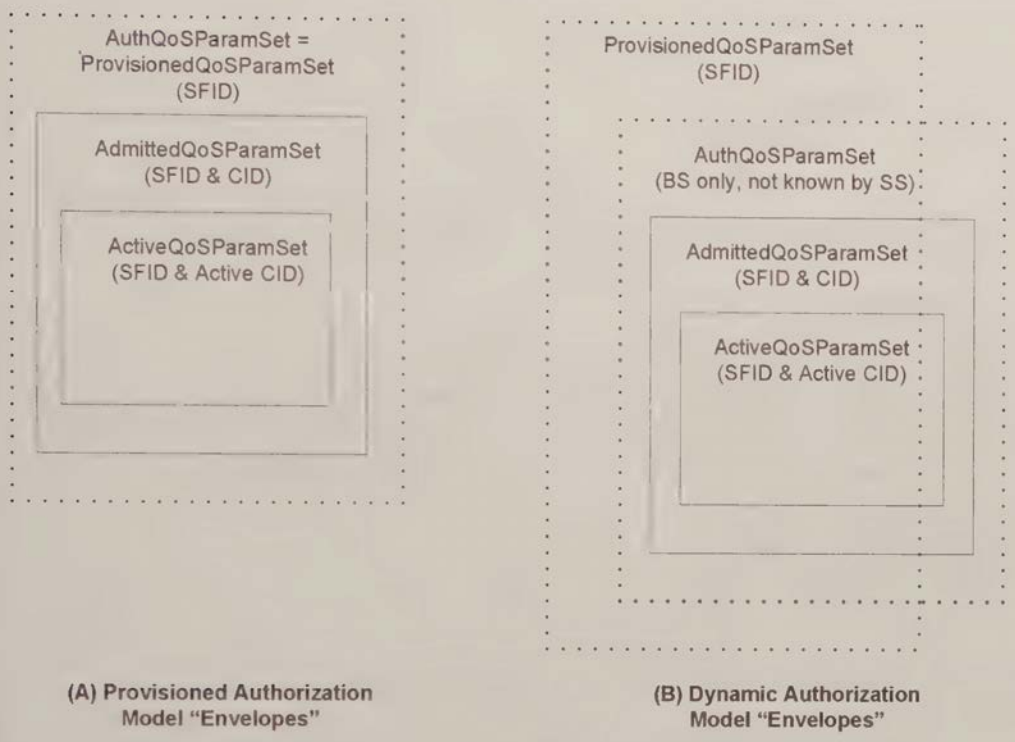


Figure 7.10 QoS Provision Model

7.5.2.3 Packet Scheduling

The packet scheduling module allocates bandwidth for connections in terms of the number of time slots allocated per connection on the TDM channel. This module also determines when a connection is allowed to transmit the data. This information is defined in the UL-MAP and DL-MAP. There are two packet scheduling modules: uplink packet scheduling and downlink packet scheduling. The downlink packet scheduling module is relatively simple because all the queues reside in the BS. Therefore, the downlink packet scheduling module can easily retrieve the state of the queues. The standard does not define the downlink packet scheduling algorithm. The downlink packet scheduling algorithms can be similar to the scheduling algorithms used in a router. The uplink packet scheduling module is more complex because the queues are distributed among multiple SSs. The uplink packet scheduling module obtains the state of the queues and the bandwidth requirements from each connection through the BW Request message as defined in the standard. The uplink packet scheduling determined by the uplink packet scheduling algorithm is reflected in the UL-MAP. Figure 7.12 illustrates the uplink scheduling process.

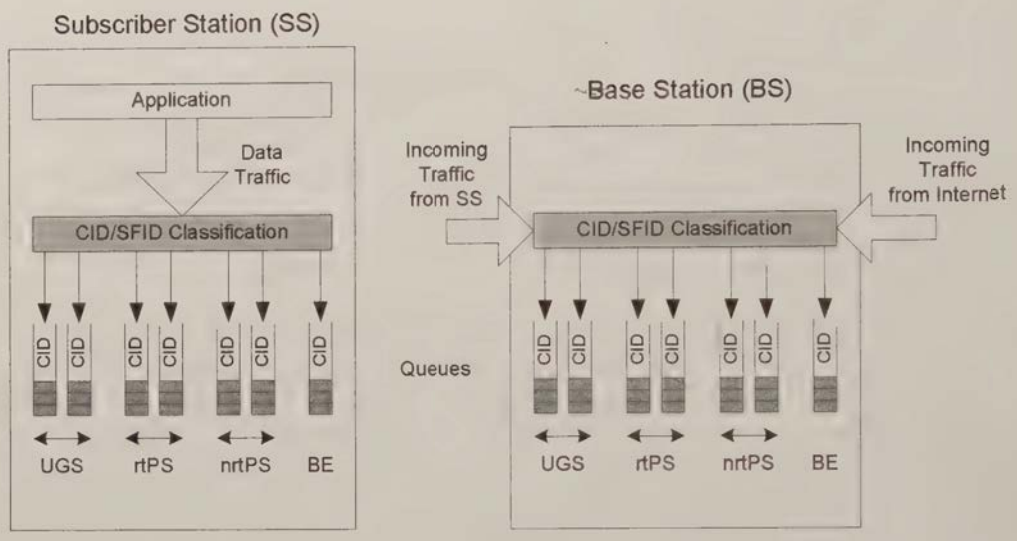


Figure 7.11 Classification Module

The standard does not define the scheduling algorithm that determines the UL-MAP. The uplink scheduling algorithm has two modes of allocating bandwidth, GPC and GPSS:

- In GPC mode, the uplink scheduling algorithm distributes the bandwidth on a per-connection basis. The intelligence of the uplink scheduling algorithm resides in the BS. The packet scheduling module at each SS is simple because it just follows the UL-MAP.
- In GPSS, the uplink packet scheduling distributes bandwidth on a per-station basis. The SS distributes its allocated bandwidth to each connection within the SS. Compared to GPC, the GPSS based uplink packet scheduling algorithm that resides at the BS is less complex. On the other hand, the packet scheduling algorithm that resides at each SS is more complex.

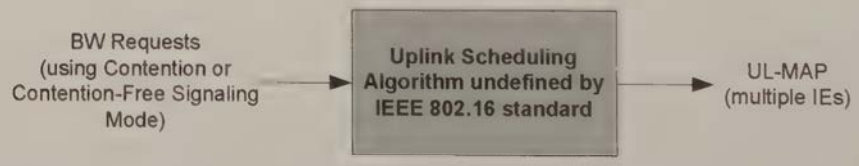


Figure 7.12 Uplink Packet Scheduling Process

Figure 7.13 summarizes IEEE 802.16 QoS architecture which can deliver multiple levels of QoS service: Quantitative Service, Qualitative Service, and BE Service.

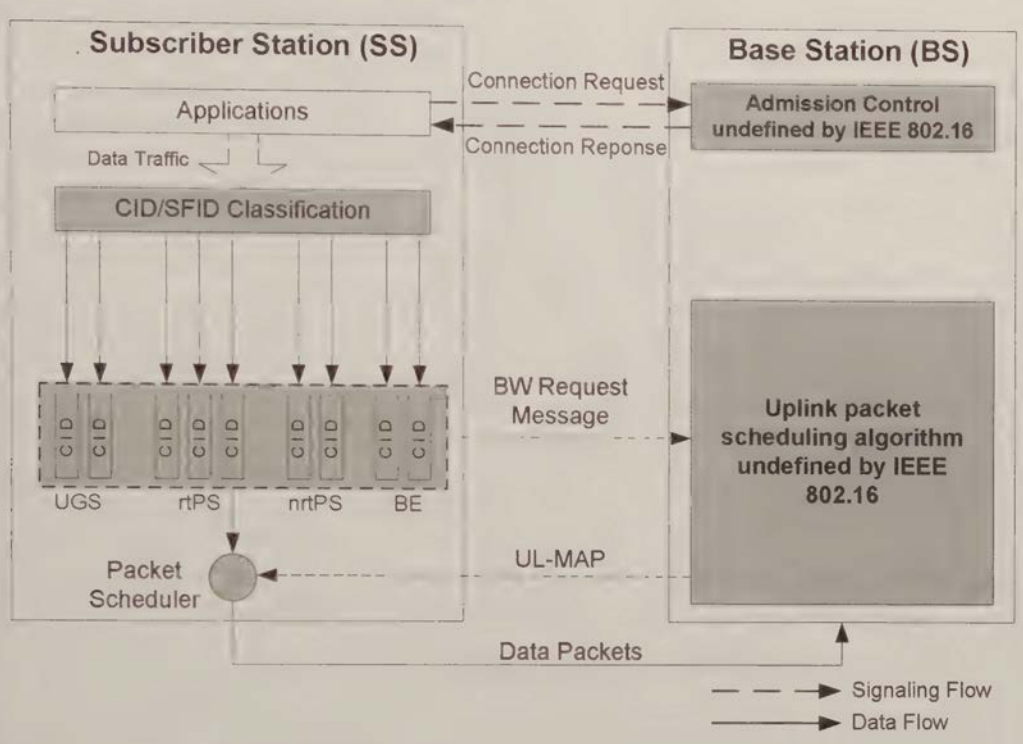


Figure 7.13 IEEE 802.16 QoS Architecture

7.6 IEEE 802.16a

The IEEE 802.16a group is considering an extension to the standard. This section provides a glimpse into the thoughts and ideas discussed in the group. The final extension, when approved, may be significantly different than these thoughts.

IEEE 802.16a focuses on extending the standard support to a mesh network topology in which direct line of sight is not required and in which SSs can communicate with each other (see Figure 7.14). This mesh topology is proposed in addition to the point to multi-point topology covered in the approved standard. Another addition is extending the physical layer to cover frequencies between 2 and 11 GHz in addition to the 10 to 66 GHz covered in the approved standard. The MAC includes the addition of an ARQ (Automatic Repeat Request) mechanism for handling communication errors as well as further signaling support for traffic scheduling algorithms that provide QoS support.

A mesh network topology is a much more complex architecture than the point to multi-point topology which extends the geographical coverage of the network. In a mesh topology, each SS can communicate with another SS, and each SS can forward traffic transmitted between the BS and the destination SS. The support of a mesh topology is more complex since we need to consider issues such as routing, collision avoidance and resolution between adjacent SSs, synchronization, and more.

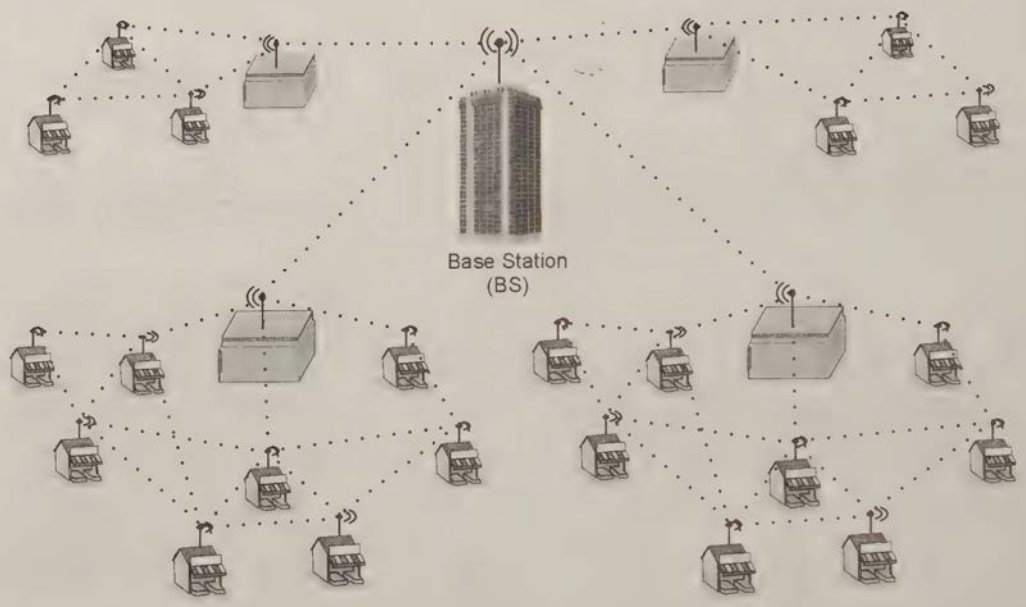


Figure 7.14 IEEE 802.16a Mesh Network Topology

The 2 to 11 GHz band considered by IEEE 802.16 provides a physical environment in which line of sight is not required. This is due to the longer wavelength as compared with the 10 to 66 GHz physical environment specified in the approved standard. The channel bandwidths used in this proposed extension are between 1.5 and 28 MHz. This addition proposes three radio technologies: single-carrier modulation format, OFDM (Orthogonal Frequency Division Multiplexing), and OFDMA (Orthogonal Frequency Division Multiple Access). OFDMA is a version of OFDM in which transmissions are carried in simultaneous subchannels compared to one channel in OFDM. OFDMA allows for increased bandwidth as well as increased resiliency to errors.

The proposal details communication in the 2 to 11 GHz frequencies where there are two license-free bands, 2.4 GHz and 5.4 GHz, that were allocated primarily for WLAN users. To avoid interfering with such users, referred to as Primary Users, the proposal defines algorithms to detect and avoid interfering with WLAN users. Referred to as Dynamic Frequency Selection (DFS), this process is an ability of a system to switch to different physical RF channels between transmit and receive activity based on channel measurement criteria. The DFS procedures test channels for primary users discontinue communication once such primary users are detected, and periodically schedule for channel testing to detect new primary users.

7.6.1 IEEE 802.16a MAC Discussions

The proposal for the mesh topology suggests both distributed and centralized scheduling mechanisms. Each mechanism can be employed with directional antennas, adaptive antenna systems, or regular omni-directional antennas. In the mesh topology, SSs that join the network do not need to interact directly with the BS but rather with their neighbor SSs.

In distributed scheduling, all the SSs need to coordinate their transmissions with their neighbors that are up to two hops away. They broadcast their available resources, requests, and grants to all of their neighbors. They can also schedule transmissions by using directed requests and grants. The proposal does not differentiate in the scheduling mechanism between downstream and upstream channels. However, SSs need to ensure that their resulting transmissions do not collide with already scheduled data and control traffic. Distributed scheduling can be executed in either a coordinated or an uncoordinated manner. The coordinated scheme employs scheduling packets that are transmitted in collision-free, regular periods within scheduling control subframes. The uncoordinated scheme allows contention based access while avoiding conflicts with the schedules established using the coordinated scheme. In case of collisions, the SS backs off and tries to retransmit.

In centralized scheduling the BS allocates all resources. The BS collects the resource requests from all the SSs that are up to a predetermined hop count from this BS. Using this information, the BS grants resources for each communication link for both downstream and upstream channels. The BS announces its decision to all of the SSs within its hop range. SSs that receive either requests from other SSs or grants from the BS forward these messages to other neighbor SSs such that all participants within the mesh are covered. To avoid multiple redundant transmissions of the same data, the mechanism assumes the existence of an algorithm where SSs know their hop count from the BS. The proposal suggests that the grant messages will not contain the actual granted schedule but rather parameters from which each SS can compute its allocated resources using a predetermined algorithm.



PART 4

Wireless Personal Area Networks





CHAPTER 8

Bluetooth

8.1 Introduction

The Bluetooth standard for wireless personal area networks is defined by the Bluetooth Special Interest Group (SIG) originally founded by Ericsson, IBM, Intel, Nokia, and Toshiba. Presently, Bluetooth SIG also includes companies such as 3Com, Agere, Microsoft, Motorola, and hundreds of other member companies. Bluetooth is named after Denmark's first king Harald Blaatand (Bluetooth in English) (940-981), son of Gorm the Old King of Denmark and Thyra Danebod. Some suggest that the king liked eating blueberries and thus his teeth became stained with the blue color. Harald Blaatand united Denmark and Norway. This inspired the name "Bluetooth," meaning "uniting" devices through Bluetooth. Bluetooth was first introduced by Ericsson as a practical technology for small form factor, inexpensive, short-range radio links between PCs, handheld devices, mobile phones, and other computing and electronic devices that communicate and interface with cellular phones and the Internet. Many companies have already introduced Bluetooth compliant products.

First we would like to compare Bluetooth with infrared technology. The Infrared Data Association (IrDA) has introduced two standards: IrDA-Data, for high-speed, short-range, line-of-sight, and point-to-point data transfer, and IrDA-Control, for lower speed communication such as wireless keyboards and joysticks. The range of IrDA is around 1 meter and requires line of sight between the transmitter and receiver. Moreover, IrDA communication is subjected to light interferences. The biggest advantage of IrDA over Bluetooth for point-to-point communication is its high throughput, which makes it suitable for multimedia applications requiring high-speed transmissions. However, Bluetooth has larger range, does not require line of sight, and is not prone to light interferences.

Bluetooth 1.1 specifies a wireless technology in the 2.4 GHz Industrial, Scientific, and Medical (ISM) band that supports a range of 10 meters. Operating on the unlicensed ISM band,

Bluetooth requires the use of spread spectrum transmission technology which is resistant to interference. Bluetooth deploys fast rate frequency hopping (compared to other Frequency Hopping Spread Spectrum (FHSS) wireless technologies) which provides robust data transmission. It supports simultaneous transmissions of both voice and data. Up to eight data devices can be connected in a cell, referred to as *piconet*. Up to 10 cells or piconets can exist within a 10 meter range. These cells are arranged in a multi-cell, referred to as *scatternet*, which increases Bluetooth's coverage range to around 100 meters. Each piconet can support up to three simultaneous full duplex voice devices. Bluetooth 1.1 standard uses a combination of circuit and packet switching. It can support an asynchronous data channel, up to three simultaneous synchronous voice channels, or a channel which simultaneously supports asynchronous data and synchronous voice. Each voice channel supports a 64 kbps synchronous voice channel in each direction. The asynchronous channel can support four asymmetric channels with data rates of up to 723 kbps and 57 kbps in the return direction, or 434 kbps for a symmetric channel.

The earlier Bluetooth 1.0 version had problems related to interoperability. For example, Bluetooth uses an FHSS radio technology in 2.4 GHz frequency band with 79-hop frequency hopping sequences. To maintain communication, the stations have to synchronize on the same hopping sequence. Some European countries allowed only 23-hop frequency hopping sequences, which created an interoperability issue. The newer Bluetooth 1.1 specification addressed this issue as well as many other interoperability issues.

Future Bluetooth 2.0 standard discusses higher speeds, improved functionality, different radios, support for ad hoc peer-to-peer networks, support for devices not currently included such as keyboards, support for high-end multimedia such as video and music, and more.

8.2 Architecture

8.2.1 Network Topology

Bluetooth defines three topologies: point-to-point (Figure 8.1), single cell (piconet) (Figure 8.2) and multi-cell (scatternet) (Figure 8.3). In a point-to-point architecture two devices communicate directly while one device becomes the master and the other becomes the slave. In a piconet, up to eight active devices communicate while one device becomes the master and the rest become slaves. Many more devices can be in "parked" mode, meaning non-active and occasionally listening to the master for synchronization and broadcast messages. The master defines the frequency hopping sequence and the clock for all the devices in the piconet. A scatternet can



Figure 8.1 Point-to-Point Topology (single slave operation)

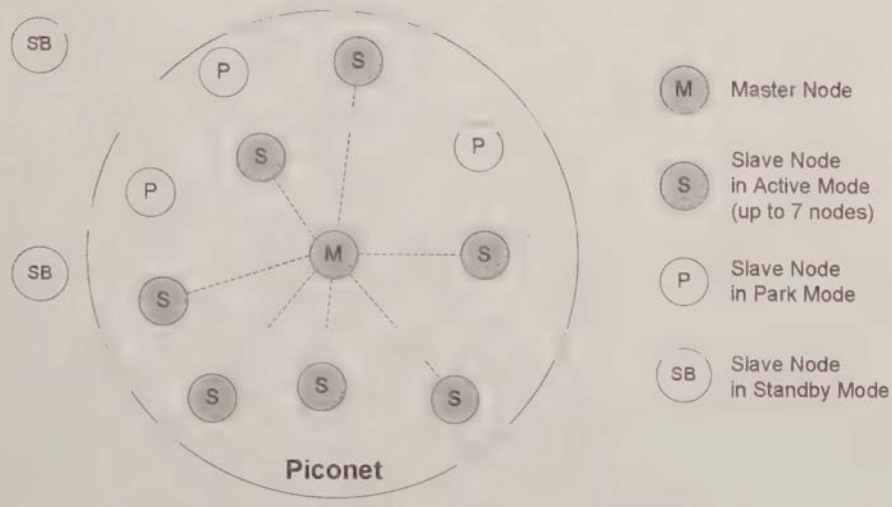


Figure 8.2 Piconet

connect up to 10 piconets. The connection between the piconets is handled by a device (either master or slave) that participates in both piconets. For example, the device can serve as a slave in both piconets or the device can serve as the master in one piconet as well as a slave in another piconet. To avoid collisions and interference between piconets, each piconet uses a different frequency hopping series which is uniquely defined by the master. However, the more piconets are active in the same area, the higher the chance that such piconets will transmit on the same frequency, leading to more collisions and interference.

8.2.2 Protocol Stack

As shown in Figure 8.4, the standard details various protocol layers: the Bluetooth Radio, the Bluetooth Baseband, the Link Manager (LM), the Logical Link Control and Adaptation Protocol (L2CAP), the Host Control Interface (HCI), the Service Discovery Protocol (SDP), the Audio/Telephony Control protocol Specification (TCS), the Radio Frequency oriented emulation of the serial communication ports (RFCOMM), the Human Interface Device (HID), TCP/IP, and other high-level protocols.

The Bluetooth Radio layer, which is parallel to the physical layer in the seven-layer Open Systems Interconnection (OSI) model, defines the requirements for a 2.4 GHz Bluetooth transceiver (i.e., frequency band, radio parameters, transmitter and receiver characteristics). The Baseband layer describes the specifications of the Bluetooth Link Controller which executes the baseband protocols and other low-level link functions. The Baseband layer includes the following modules: LM, HCI, and L2CAP. The LM protocol specifies the link setup and control. The HCI provides a command interface to the Baseband Link Controller and LM and access to the hardware status and control registers. The L2CAP supports higher level protocol multiplexing,

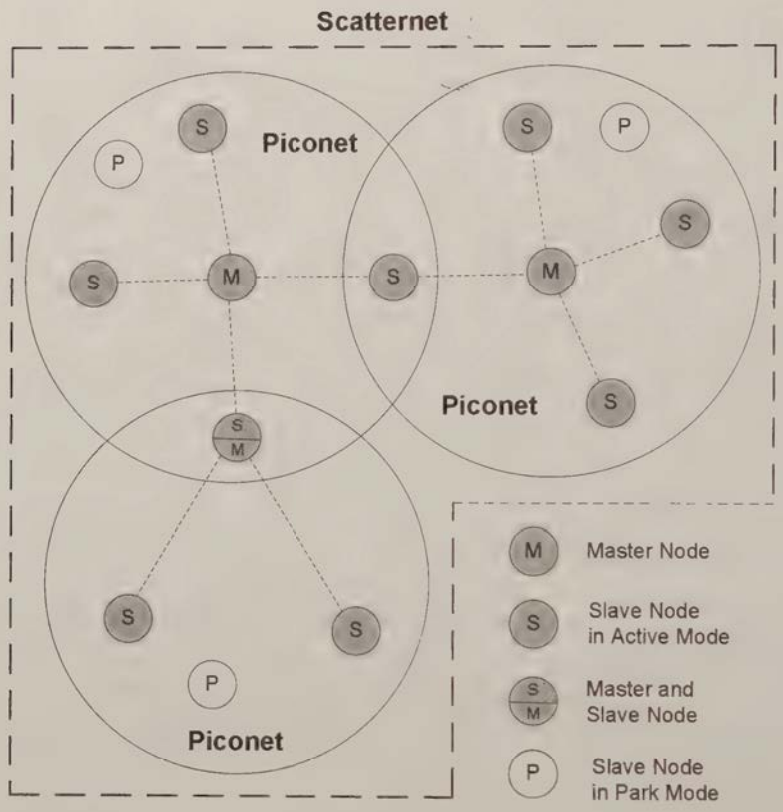


Figure 8.3 Scatternet

packet segmentation, and packet reassembly, and it conveys QoS information. The RFCOMM protocol provides emulation of serial ports over the L2CAP protocol. The Service Discovery Protocol (SDP) provides ways for applications to discover which services are available.

8.3 Physical Layer

The Bluetooth radio uses FHSS in 79 frequencies, each in a 1 MHz range from 2.402 GHz to 2.480 GHz. The radio transceiver changes the hopping frequency in a pseudorandom fashion determined by the master. The hopping rate is 1600 hops per second. GFSK (Gaussian Frequency Shift Keying) is used for the modulation scheme. Each device is classified into power classes 1, 2, and 3. Power class 1 is for long-range transmission of around 100 meters. Power class 2 is for short-range transmission of around 10 meters with less power output than power class 1. Power class 3 is for short-range transmission of around 10 cm, using minimal transmission power. The power control function regulates the transmission power for optimizing the power consumption and channel interference level.

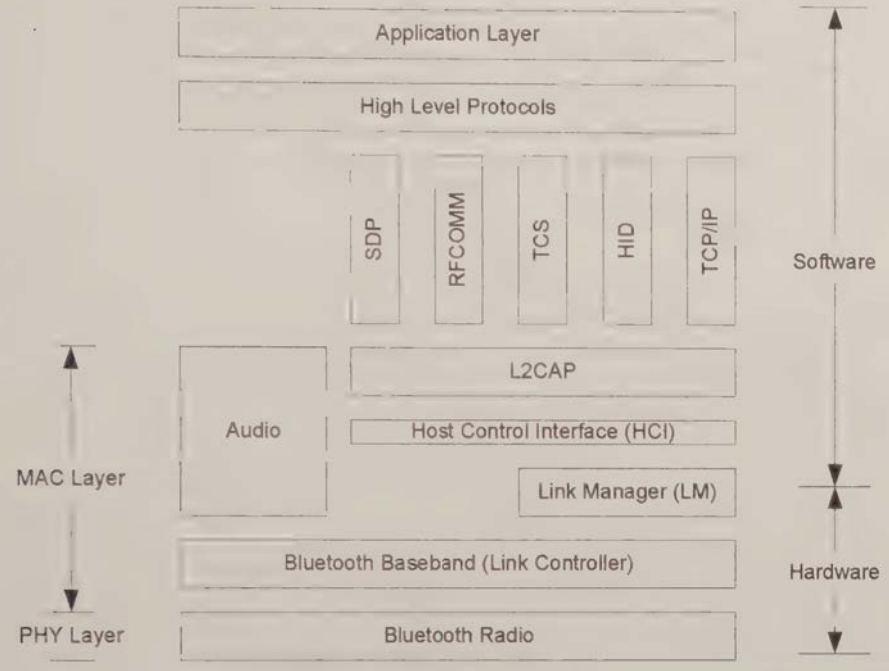


Figure 8.4 Bluetooth Protocol Stack

8.4 Bluetooth Baseband

The Baseband executes the key functions of the Media Access Control (MAC) layer and Data Link Control (DLC) layer. The MAC functions include channel access and packet handling through asynchronous and synchronous links. The link control protocol of the Baseband is implemented as a Link Controller, which works with the LM for carrying out link level routines such as link connection and power control.

8.4.1 Channel Definition and Channel Access

A piconet consists of a master node, active slave nodes (up to seven nodes), and several inactive slave nodes (up to 200 nodes). All connections (or links) in the piconet take place only between the master and slaves in point-to-point or point-to-multipoint fashion. There is no direct connection between slaves. The communication channel is structured into time slots, each with duration of 625 μs. The master uniquely defines the frequency hopping sequence based on its device address (48 bits). Generally, all devices hop or change the frequency every time slot (with max hopping rate of 1600 hops per second, which corresponds to 625 μs time slot duration) except in the case of multi-slot packet transmission, which we will describe later in this section. All slaves (both active and inactive) in a piconet are required to synchronize the clock and frequency hopping sequence with the master. A packet transmission starts at the beginning of a time slot and can last for one, three, or five time slots.

Bluetooth deploys a full duplex, Time Division Duplex (TDD) protocol in which transmission and reception alternate. By alternating time slots, the master and slave can transmit packets according to the TDD protocol. The master starts its transmission only in even numbered time slots and the slave starts its transmission only in odd numbered time slots. Figure 8.5 illustrates a single time slot packet transmission.

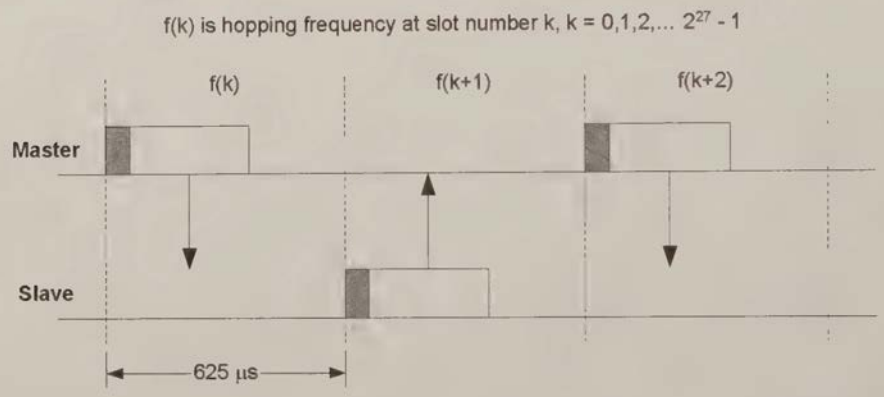


Figure 8.5 Example of a Single Time Slot Transmission Using TDD

A single-slot packet is transmitted with a hopping frequency of the current Bluetooth clock value. In case of a multi-slot packet, the packet will be transmitted with the hopping frequency of the first time slot for the entire duration of packet transmission. The multiple time slot transmission is shown in Figure 8.6.

Bluetooth aims to support both voice and data transmission by using a combination of both circuit and packet switching protocols. Bluetooth Baseband defines two types of transmission links between a master and a slave: synchronous connection-oriented (SCO) link and asynchronous connectionless (ACL) link. SCO is used for synchronous packet transmission while ACL is used for asynchronous data transmission.

8.4.1.1 SCO and ACL links

The SCO link is a symmetric, point-to-point link between the master and a single slave in the piconet. The SCO link is established by the master by sending an SCO setup packet via the LM protocol. This packet contains timing parameters such as reserved time slots. The master maintains the SCO link by reserving time slots at regular intervals—that is, establishing a circuit switched connection. A slave can support up to three SCO links from the same master or two SCO links if the links originate from different masters. The master can support up to three simultaneous SCO links to the same slave or to different slaves. With the reserved time slots at regular intervals, the SCO link provides bounded delay channel access which is suitable for time sensitive applications such as voice. The master sends packets, referred as SCO packets, at regu-

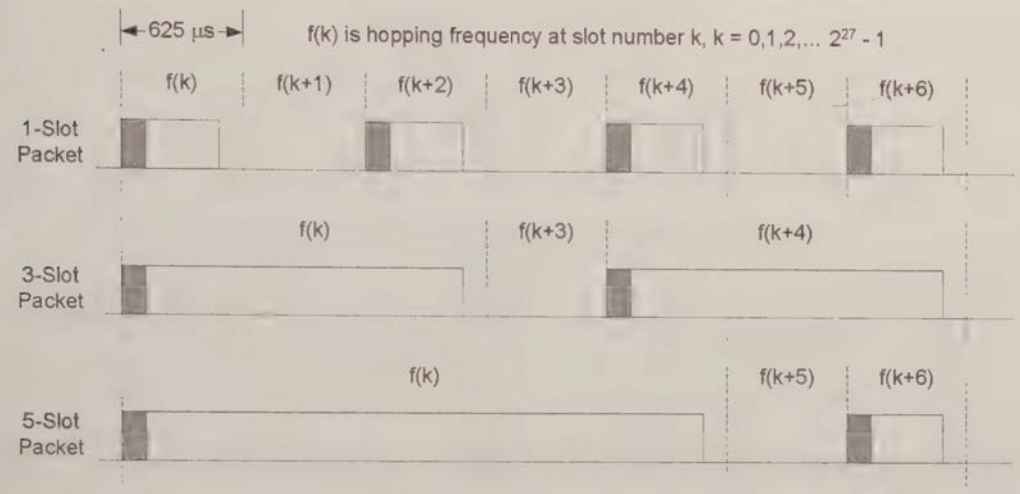


Figure 8.6 Multi-Slot Packets

lar intervals, called SCO intervals and counted in slots in the reserved master-to-slave slots. The SCO slave is allowed to respond with an SCO packet in the following reserved slave-to-master slot. Figure 8.7 shows two SCO links with different SCO intervals. Bandwidth allocation to SCO links and channel access delay are determined by the SCO interval.

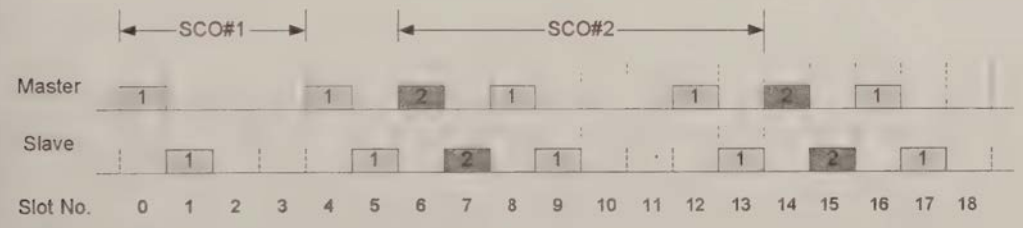


Figure 8.7 SCO Links

There is no packet retransmission in SCO transmissions. The error handling relies on the forward error correcting (FEC) scheme and the codec used. SCO is designed to be used by voice packet transmissions. Bluetooth specification categorizes the voice packet into three types: HV1, HV2, and HV3. HV stands for high-quality voice. The differences between the three types of HV are the number of information bytes and the error correcting scheme. HV1 provides the lowest number of information bytes (low data rate) but uses the most robust error correction scheme (i.e., 1/3 FEC). HV3 provides the highest number of information bytes (high data rate) but does not use any error correction scheme. While achieving high data rate, HV3 is susceptible to channel interference. Error handling for HV3 only relies on the codec of voice content.

The ACL link is a point-to-multipoint link between the master and all the slaves in the piconet. In the slots not reserved for SCO packets, the master can establish an ACL link on a per-slot basis (i.e., packet switched connection) to any slaves, including the slave(s) already engaged in an SCO link. Only a single ACL link between a master and a slave can exist. That means that multiple packet switched connections on a slave are aggregated to the same single ACL link. ACL supports both asynchronous and isochronous services. For most ACL packets, packet retransmission is supported to ensure integrity. A slave can respond to an ACL packet in the slave-to-master slot only if it has been addressed in the preceding master-to-slave slot. ACL packets that are not addressed to a specific slave are considered as broadcast packets. The slave's transmission of an ACL packet is controlled by a POLL packet sent from the master in the master-to-slave slot. When receiving the POLL packet, the slave responds by sending data packets in case it has packets to send, or in case it has nothing to send, it sends a NULL packet in the following slave-to-master slot. The bandwidth allocation for a slave's ACL link is controlled by the polling frequency of the master. The more frequently a slave receives poll packets, the more bandwidth a slave is allocated. Bluetooth specifications define various types of asynchronous data packets for the ACL link (i.e., DM1, DH1, DM3, DH3, DM5, DH5, AUX) which have different payload size, error detection scheme, error correcting scheme, and achievable data rate. The ACL link is used by asynchronous data packets and control packets (i.e., from LC, LM, L2ACP).

8.4.1.2 Logical Channels

Bluetooth defines five logical channels used to transfer different types of information:

- *Link Control (LC) Channel:* This channel carries low-level link control information such as flow control and payload characterization.
- *Link Manager (LM) Channel:* Typically it carries control information exchanged between the master's LMs and the slaves.
- *User Asynchronous (UA) Channel:* It carries L2CAP transparent asynchronous user data. These data may be transmitted in one or more baseband packets.
- *User Isochronous (UI) Channel:* It carries L2CAP transparent user isochronous data. These data may be transmitted in one or more baseband packets.
- *User Synchronous (US) Channel:* It carries transparent synchronous user data. This channel is carried over the SCO link.

LC and LM are used at the link control level and LM. UA, UI, and US are used to carry user information. The information on the LC channel is carried in the packet header, while the information for all other channels is carried in the packet payload. The information for the US channel is carried only by the SCO link. The information for the UA and UI channels is normally carried by the ACL link; however, it can also be carried by the data portion of the combined data-voice packet on the SCO link. The information on the LM channel can be carried either by the SCO or the ACL link. Figure 8.8 shows the flow of each logical channel.

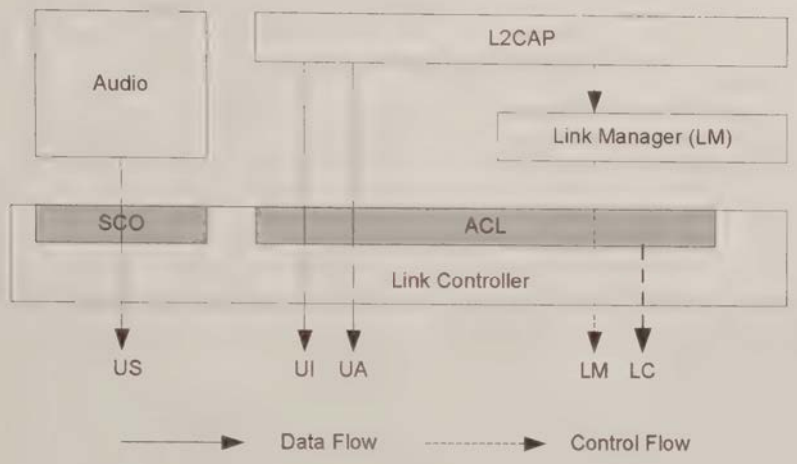


Figure 8.8 Logical Channels

8.4.1.3 Packet Format

Figure 8.9 shows the standard packet format, which includes the Access Code, LC header, and payload.

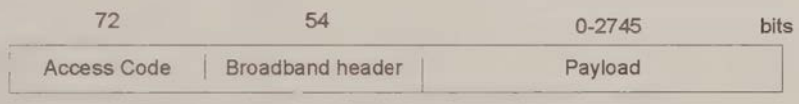


Figure 8.9 Standard Packet Format

There are three types of access codes:

- *Channel Access Code (CAC):* CAC is used to identify a piconet. CAC is determined by the device address of the piconet's master. Therefore, CAC is unique in each piconet. CAC is normally included in the user data packet.
- *Device Access Code (DAC):* DAC is used to identify a device. DAC is uniquely determined by the device address. DAC is used for the paging procedures. For example, in the paging procedure, the master will send the DAC of the slave with which it wants to establish a connection. Then, the slave responds back with its DAC. The DAC is included in the paging and paging response messages which include only the access code field (no header and payload).
- *Inquiry Access Code (IAC):* IAC is used for the inquiry procedures. IAC is included in the inquiry message which includes only the access code field (no header and payload). The master sends inquiry messages to collect the device addresses of slaves in the transmission range.

Figure 8.10 shows the packet format of the ACL and SCO links. An SCO packet contains the voice payload while an ACL packet contains the data payload. Some SCO packets can also contain data payload. The LC can use either SCO or ACL packet with or without payload (i.e., NULL and POLL packets do not have data payload while frequency hop synchronization (FHS) packets include control information in the data payload). The LM includes its control information in the data payload.

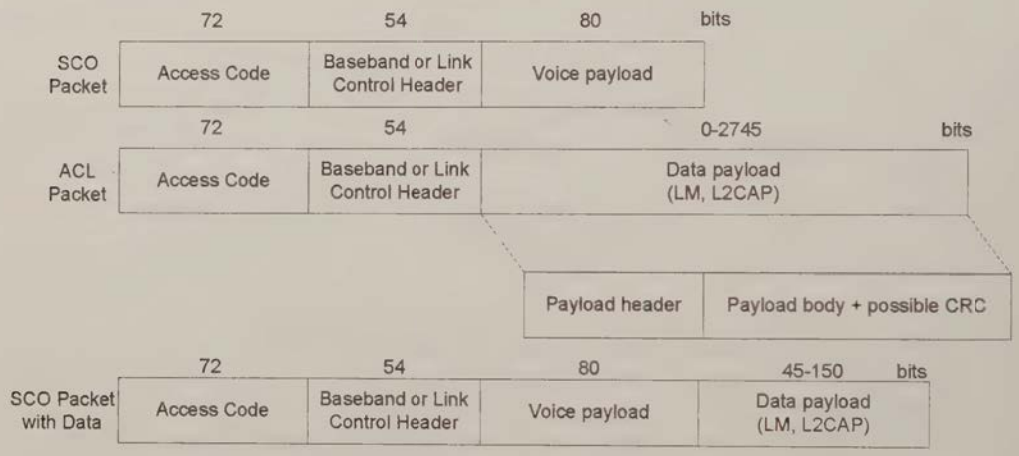


Figure 8.10 SCO and ACL Packet Format

8.4.2 Link Control

Before the master and slaves can communicate with each other, they have to establish the piconet using link control mechanisms based on link control states. Link control information is included in the link control header of the packet as shown in Figure 8.10.

8.4.2.1 Link Control States

A station can be in one of a number of link control states (see Figure 8.11). There are two major states: Standby and Connection. In addition, there are substates with their associated procedures: page, page scan, inquiry, and inquiry scan. These procedures are used to establish a piconet or add new slaves to a piconet. The substates are transient states between the Standby and Connection states.

8.4.2.1.1 Standby State The station default state is Standby state, which is a low-power consumption mode. The Standby stations are not associated with any piconets. Occasionally, the Link Controller may leave this state to scan for page or inquiry messages, or to page or inquiry itself. If the station received such a message (i.e., page message, inquiry message), it will enter the Connection state and become a slave. If the station itself transmits a successful page message or inquiry message, then it will enter the Connection state and become the master.

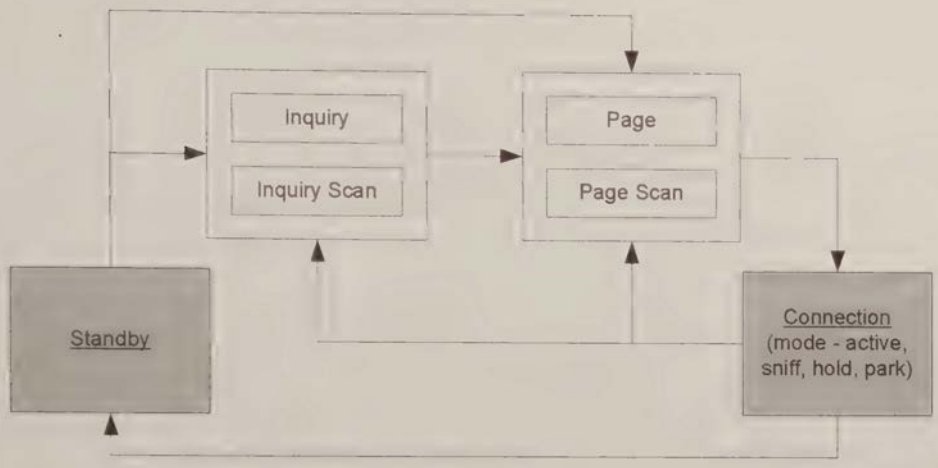


Figure 8.11 Link Control State Diagram

Inquiry procedures (see Figure 8.12) are the interaction processes between the station in the inquiry substate (called inquiry station) and the station in the inquiry scan substate (inquiry-scan station). Inquiry is used where the destination's device address is unknown to the source. The inquiry stations are the stations that want to discover which new stations are within their transmission range. The inquiry-scan stations are the stations that want to be discovered by other stations. The inquiry station transmits the inquiry message (containing IAC) with a frequency

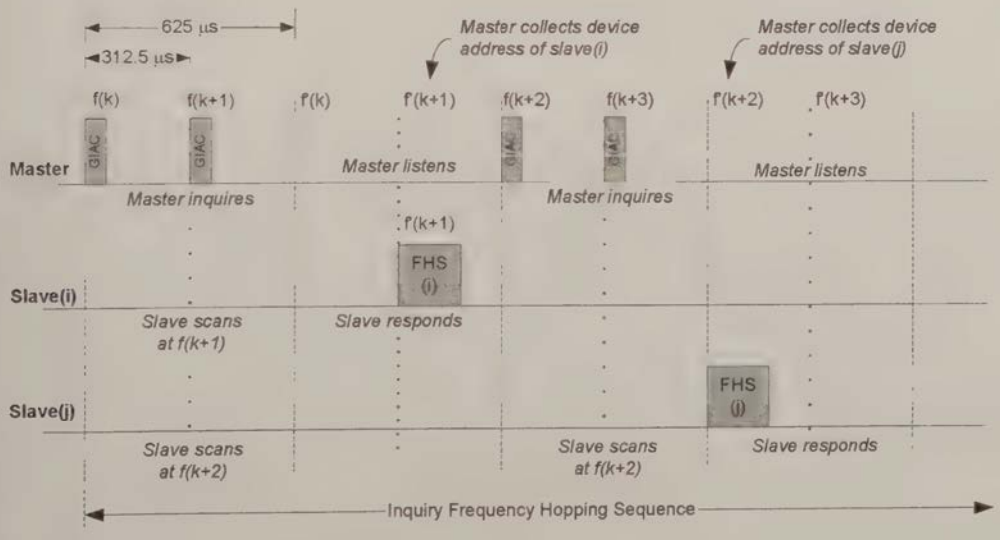


Figure 8.12 Inquiry Sequences

that follows the inquiry frequency hopping sequence. The inquiry-scan station scans the inquiry message at a single hop frequency. The inquiry-scan stations can receive the inquiry message whenever the inquiry station transmits at the same frequency. After receiving the inquiry message successfully, the inquiry-scan station will send to the inquiry station the response message that contains the device address and the clock. In summary, during this process, the inquiry station collects the device addresses and clocks of all stations that respond to the inquiry message. Based on this information the inquiry station can issue a page message to establish connection to known stations.

Paging procedures are used to establish new connections between the master and a slave. The paging substate is deployed by the master to establish the new connection with the slave which is in the paging-scan substate. The paging procedures are similar to the inquiry procedures. The master tries to connect to the targeted slave by repeatedly transmitting the slave's device access code (DAC) in several hop channels. Since the clocks of the master and the slave are not yet synchronized, the master does not know exactly when the slave wakes up and on which hop channel. The master predicts the hop channels where the slave should be, for example, by recalling details of the last exchange of packets. Based on this prediction, it transmits a series of the same DACs at the predicted channel as well as other channels and listens between transmissions for a response from the slave. In Page Scan, a slave listens for its own DAC for the duration of the scan window. During this window, the slave listens at a single hop frequency, its correlator matched to its device access code. When a slave enters Page Scan, it selects the scan frequency according to the page hopping sequence corresponding to this station's device address. Figure 8.13 depicts the paging sequences between the master and a slave. The master transmits paging messages more frequently (3200 packets per second or once every 312.5 μs) than the normal data packet transmissions (1600 packets per second or once every 625 μs). The master determines its paging frequency hopping sequence based on its device address. The slaves

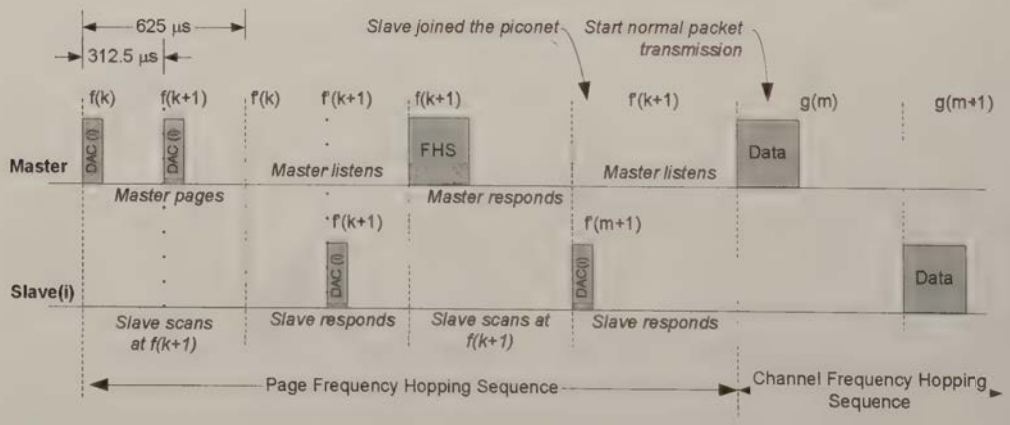


Figure 8.13 Paging Sequences

listen or scan at a fixed frequency. If a slave receives the paging message, it will transmit a paging response message which contains its DAC. Then, the master transmits the FHS message in order for the slave to synchronize 1) to the channel frequency hopping sequence and 2) to the clock. Then, the slave replies with a FHS packet and becomes an active node of the piconet.

In Standby, no connection has been established and the station can use all of its capacity to support Page Scan. If desired, the station can place ACL connections in the Hold mode or even use the Park mode before entering Page Scan. SCO connections are preferably not interrupted by Page Scan. In this case, Page Scan may be interrupted by the reserved SCO slots which have higher priority than Page Scan.

8.4.2.1.2 Connection State At the connection state, a Bluetooth station becomes the member of a piconet. The station can be in one of the following four connection modes: Active, Sniff, Hold, and Park. These connection modes are used to save power and to allow stations to communicate with different piconets.

Active Mode: In this mode the station can actively communicate on the wireless channel. A piconet is allowed to have up to eight active mode stations at the same time. The master schedules the transmission based on traffic requirements to and from the slaves. In addition, the master supports regular transmissions to keep slaves synchronized to the channel. Active slaves listen during the master-to-slave slots. If an active slave is not addressed, it may sleep until the next new master transmission. A periodic master transmission is required to keep the slaves synchronized to the channel. Since the slaves only need the channel access code for synchronization, any packet type can be used for this purpose.

Sniff Mode: Slave stations can change to a power-saving mode in which the station's activity is decreased. In this Sniff mode the slave station listens to the piconet less often, thus reducing power consumption. If a slave participates on an ACL link, it has to listen to the master traffic every ACL slot. The time interval between consecutive listening events is programmable by the application.

Hold Mode: Slave stations can change into another power-saving mode. During the Connection state, the ACL link to a slave can be in Hold mode. This means that the slave temporarily does not support ACL packets while continuing to support possible SCO links. While in Hold mode, the station can execute procedures such as scan, page, inquire, or communicate with another piconet. Before the slave enters Hold mode, the master and slave have to agree on the time duration that the slave can remain in this mode. Once this time expires, the station will return to Active mode, synchronize to the traffic, and wait for instructions from the master.

Park Mode: In this mode the slave station is still synchronized to the piconet but does not communicate. Parked stations give up their active member address and only occasionally listen to the traffic from the master to synchronize and check for broadcast messages. To support parked slaves, the master establishes a beacon channel where one or more slaves are parked. The master transmits on this channel at constant time intervals a series of beacon slots.

8.4.2.2 Piconet Establishment and Operation

The piconet is governed entirely by its master. The master's address determines the frequency hopping sequence and the channel access code (the code that uniquely identifies the piconet). Since Bluetooth is based on TDD, every Bluetooth station has an internal system clock which determines the timing and hopping of the radio transceiver. This clock has no relation to the time of day and can therefore be initialized at any value, does not require adjustments, and is never turned off. For synchronization with other units, only offsets are used and added to the native internal clock. To ensure synchronization, the master sets the timing of all slaves based on its clock by transmitting to the slaves its clock reading. The slaves add an offset value to their native clocks so they can be synchronized to the master clock. Since the clocks are free-running, the offsets have to be updated regularly. Establishing piconet is controlled by link control state. Figure 8.14 shows the piconet establishment sequence.

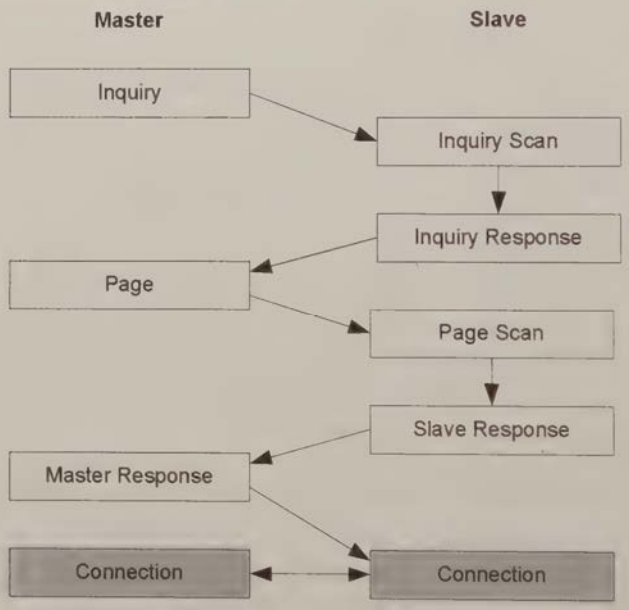


Figure 8.14 Piconet Establishment Sequences

First, the master executes the inquiry process to obtain the device addresses and clocks of the slaves within the transmission coverage. Then, the master performs the paging process to establish the connections with the specific slave. Finally, both the master and the active slaves can communicate with each other.

8.4.2.3 Scatternet Establishment and Operation

To increase the coverage area and the number of supported stations, several piconets may coexist in the same area. Each piconet's master specifies a different hopping series. In addition, the packets carried on the channels are preceded by different channel access codes as determined by the master addresses. As the number of coexisting piconets increases, the probability of collisions increases since different piconets may transmit on one of the 79 frequency channels. If several piconets coexist in the same area, a station can participate in two or more piconets by applying time multiplexing. In each piconet, the station will use this piconet hopping series, master device address, and proper clock offset. A Bluetooth station can act as a slave in several piconets but as a master in only one piconet. A group of connected piconets is referred to as a scatternet.

Stations must use time multiplexing to switch between the piconets in which they participate. In case of ACL links, a station can request to enter the Hold or Park mode so that during this mode it can join the other piconet. Units in the Sniff mode may have sufficient time to visit another piconet between the sniff slots. If SCO links are established, other piconets can only be visited in the non-reserved slots between the sniff slots. Since coexisting piconets do not have a mechanism to synchronize clocks, additional guard time should be considered for proper switching.

A master can become a slave in another piconet by being paged by the master of this other piconet. On the other hand, a slave participating in one piconet can be a master in another piconet. Since the paging unit always starts out as master, a master-slave role exchange will be required. Such an exchange may involve a reorganization of piconets since slaves may change from the current master to the new master. This is a complex process since it requires all stations to resynchronize their clocks, transit to the frequency hopping series of the new masters, and adjust to the new master device addresses. This reorganization of piconets should be done without interfering with ongoing communication activities and without involving the users. The option of setting up a completely new set of piconets will require a long period of time. Another option is to have the new masters utilize their knowledge of timing and hopping sequences to reduce the time for such an exchange. Because of the complexity of the piconet reorganization process, it is currently left outside the scope of the Bluetooth standard. Therefore, Bluetooth developers need to design and implement their own exchange algorithm.

8.4.2.4 Link Supervision

Several events can cause stations to loose their communication link. Such events can be power failure, movement out of range, and interference from outside sources. Hence, the standard defines a process in which the link is monitored by the master and the slave using link supervision timers for both SCO and ACL connections. When a station receives a packet with a valid slave address, this timer is reset. However, when this timer reaches a predetermined value when the station is in the Connection state, the connection is reset. This timer timeout period is negotiated at the Link Management (LM) such that it is longer than Hold and Sniff periods. If the slave is in Park mode, then link supervision is done by un-parking and re-parking the slave.

8.5 Link Manager (LM)

The Link Manager is responsible for link setup, security, authentication, link configuration, timer setup, and other control procedures (Figure 8.15). The communication between the stations' LM modules is done using the Link Manager Protocol (LMP). The LM sends its control message through Data Medium rate (DM1) or Data Voice (DV) packets. Due to the fact that DM1 and DV packets use reliable links (using 2/3 FEC and CRC), there are no explicit acknowledgment messages. LM messages have higher priority than user data.

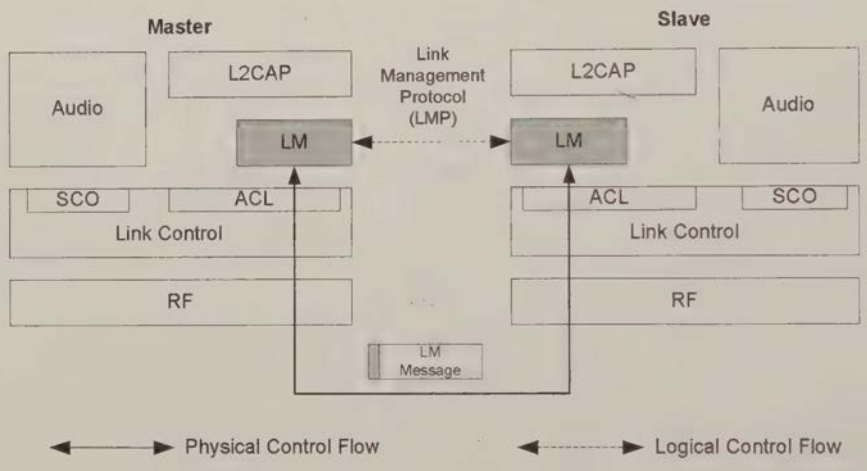


Figure 8.15 Link Manager

LM supports upper layer applications that execute control algorithms such as managing the connection states (Park, Hold, Sniff, Active), establishing security, and supporting QoS. These control algorithms may be self contained or may allow input from users. We will discuss more about the QoS aspects of the LM protocols in a later section.

8.6 Host Control Interface (HCI)

As we showed in Figure 8.4, which describes the Bluetooth protocol stack, HCI is the layer between L2CAP and LM. The HCI provides a standard interface to Bluetooth so that upper layers can be independent and transparent from the host hardware implementation. It provides interface to the LM and access to the hardware status and control registers. HCI consists of multiple parts (i.e., HCI driver, HCI firmware) that reside in both Bluetooth Host and Bluetooth hardware. Figure 8.16 shows HCI relationship to the Bluetooth system.

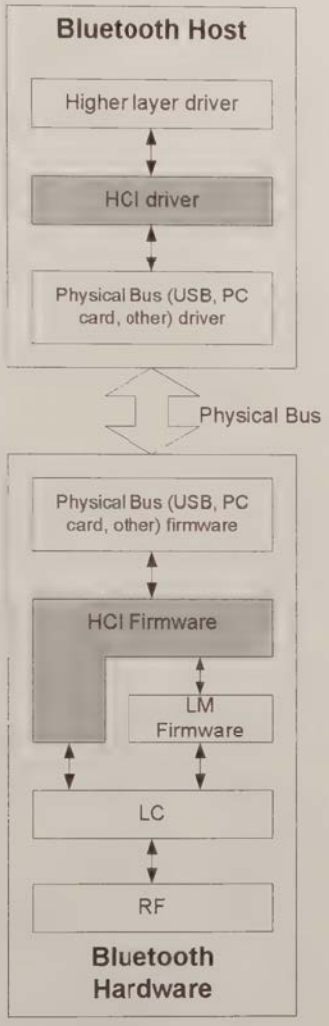


Figure 8.16 HCI Relationship to the Bluetooth System

8.7 Logical Link Control and Adaptation Protocol (L2CAP)

The Logical Link Control and Adaptation Layer Protocol (L2CAP) provides connection-oriented and connectionless data services to upper layer protocols. The L2CAP specification is defined for only Asynchronous Connectionless (ACL) links. Support for Synchronous Connection-Oriented (SCO) links, mainly used for real-time voice traffic with reserved bandwidth, is not yet provided in L2CAP. Bluetooth recommends that L2CAP should include simple and low-overhead programs so they can fit in devices with limited computational, power, and memory resources.

These data services include QoS support, group support, and protocol multiplexing, segmentation and reassembly. Protocol multiplexing enables an application to simultaneously use several higher layer protocols such as TCP/IP and RFCOMM.

L2CAP labels each connection by a channel identifier (CID) and assumes that each channel is a full duplex connection and may have a QoS flow specification. The L2CAP connection establishment process allows exchange of the expected QoS between two Bluetooth stations. Moreover, each station monitors the resources used to ensure that the required QoS contract is provided. Again, these QoS algorithms are left to the developer and are not provided as part of Bluetooth.

L2CAP group support permits higher level applications to map groups (group of addresses) to piconets. The Baseband layer supports the concept of a piconet for a unique group. Without group support, higher level applications would need to establish a complex communication process directly with the Baseband Protocol and LM layers.

8.8 Higher Bluetooth Layers

8.8.1 RFCOMM

RFCOMM enables compatibility with applications that use the serial port as their main communication bus. RFCOMM conveys all of the RS232 control signals and supports remote port configuration.

8.8.2 Service Discovery Protocol (SDP)

The service discovery protocol (SDP) provides tools for applications to discover which services are available and to determine the nature of these services.

8.8.3 Audio and Telephony Control (Tel Ctrl)

This layer includes the interface needed to connect and disconnect a telephone call, including signaling the devices that participate in the connection. Telephony audio links are established with synchronous links and therefore do not go through the same L2CAP-to-LM path that asynchronous links go through. In other words, audio links may be thought of as direct Baseband-to-Baseband links.

8.9 Profiles

Bluetooth profiles specify a set of basic standards to ensure device interoperability. In Bluetooth Specifications, v. 1.1, there are 13 Bluetooth profiles. They include profiles such as Generic Access, Cordless Telephony, Intercom, Serial Port, Headset, Dial-up Networking, Fax, and LAN Access and Synchronization (exchange of information between applications such as calendar).

8.10 QoS Support

Bluetooth aims to support both synchronous and asynchronous data services. The key concept introduced by Bluetooth for QoS support is its definition of two data link types, SCO and ACL. Furthermore, Bluetooth also defines QoS as enabling MAC and QoS signaling. In this section, we take a closer look into Bluetooth's QoS mechanisms. Figure 8.17 shows Bluetooth's QoS architecture.

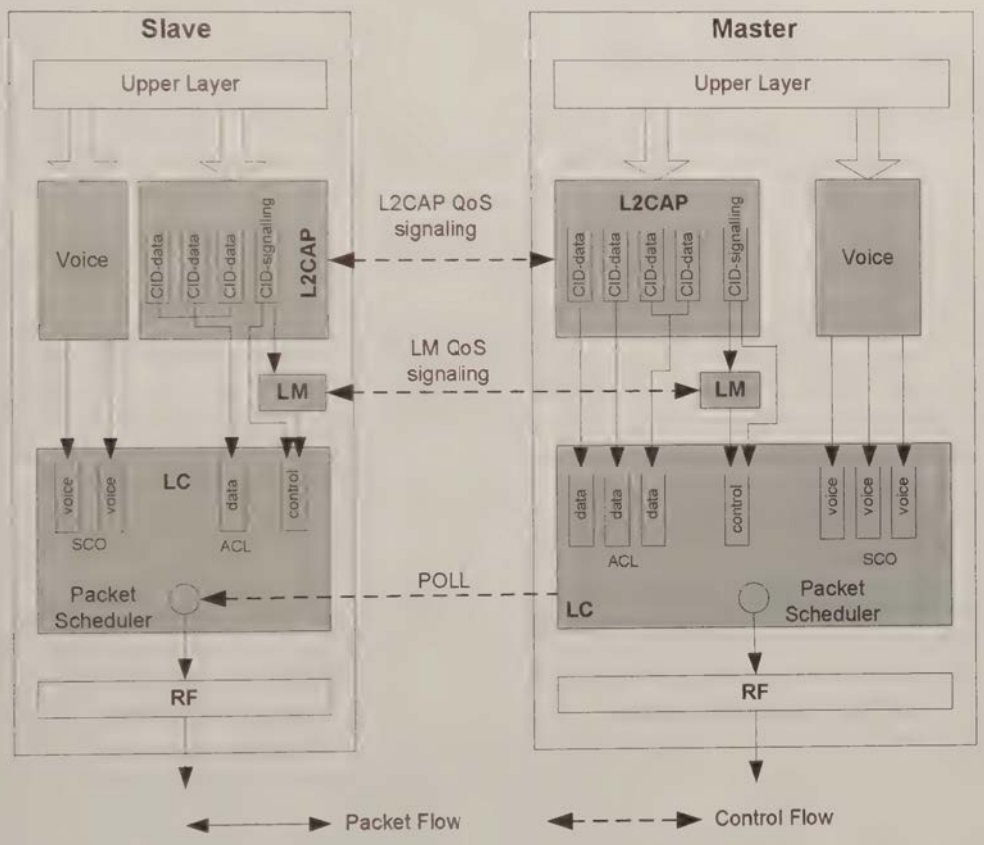


Figure 8.17 Bluetooth QoS Architecture

8.10.1 Classification

Bluetooth defines packet classification at the L2CAP layer. All asynchronous data packets are assigned a CID as shown in Table 8.1.

Table 8.1 Channel Identifiers

CID	Description
0x0000	Null identifier
0x0001	Signaling channel
0x0002	Connectionless reception channel
0x0003 – 0x003F	Reserved
0x0040 – 0xFFFF	Dynamically allocated

For a connection-oriented application, CIDs are dynamically assigned on both end stations. For a connectionless application, a fixed CID (0x0002) is assigned to the receiver and a dynamic CID is allocated to the sender. For L2CAP signaling, both end stations are allocated CID=0x0001. A classifier will use the CID to identify the packets and forward them to the appropriate queue. Bluetooth does not define how to classify the voice packets that do not go through the L2CAP layer.

8.10.2 Channel Access

As described in Section 8.4, Bluetooth deploys a TDD channel access. The master is the only station that transmits on the downlink direction (master-to-slave time slot). On the other hand, the uplink (slave-to-master time slot) is shared by multiple slave stations. The uplink channel access is controlled by the master which allocates reserved time slots for SCO packets and uses POLL control packets for ACO packets. After receiving a POLL, the slave is allowed to transmit ACO packets. Bluetooth uses a collision-free channel access scheme which provides tight channel access control. Therefore, these schemes can provide QoS support for applications with strict QoS requirements.

8.10.3 Packet Scheduling

We differentiate between two packet scheduling schemes: Intrastation packet scheduling and Interstation packet scheduling. There are three types of packets: SCO packets, ACL packets, and Control packets (which use the ACL link). In the LC module of the slave, a slave can have up to three SCO links for the same master or up to two SCO links for different masters, one ACL link and one control link. In the LC module of the master, the master can have up to three SCO links,

one ACL link for each slave, and one control link. Using these packet scheduling mechanisms, Bluetooth can provide strict QoS service for voice applications using the SCO link and loose QoS service for asynchronous applications using the ACL link.

8.10.3.1 Intrastation Packet Scheduling

SCO packets are transmitted in the reserved time slot and ACL packets are transmitted into the rest of the time slots. Therefore, Bluetooth implicitly deploys strict priority packet scheduling. SCO has higher priority than ACL. Furthermore, Bluetooth also explicitly defines that control packets have higher priority than ACL packets. In summary, SCO has the highest priority followed by control packets and ACL packets. There is only one ACL link in a slave. Asynchronous data packets from different applications initiated by the same slave will be aggregated in the same queue. All packets experience the same QoS. Moreover, it is possible that an asynchronous application can abuse the other asynchronous applications by using the bandwidth as much as it can.

8.10.3.2 Interstation Packet Scheduling

The slave is allowed to transmit in the reserved time slot (for SCO) or in response to a POLL message (for ACL). The bandwidth allocation for SCO packets depends on the number of reserved time slots which are assigned by the signaling process. The bandwidth allocation for ACL packets depends on the number of POLL packets received (polling frequency) which is controlled by the master. Bluetooth does not define the polling algorithm which determines the polling sequence and polling frequency.

8.10.4 QoS Signaling

There are two levels of QoS signaling: L2CAP QoS signaling and LM QoS signaling.

8.10.4.1 L2CAP QoS Signaling

L2CAP QoS signaling exchanges the QoS information between L2CAP of master and slave. Such QoS information indicates the traffic characteristic and QoS requirements of applications (called FlowSpec) such as Token Rate, Token Bucket Size, Peak Bandwidth, and Latency and Delay variation. The QoS information is included in the Configuration Parameter Option field of the L2CAP control message. L2CA_ConfigReq is the L2CAP control message that performs the connection request. L2CA_ConfigReq requests the traffic characteristics and QoS requirement of outgoing data flow to the master. Then the master sends L2CA_ConfigRsp in response to L2CA_ConfigReq. L2CA_ConfigRsp indicates the accepted traffic characteristics of the incoming data flow. Bluetooth does not define the admission control process that decides the accepted traffic characteristics. Bluetooth does not define the algorithm that maps the FlowSpec to polling frequency and polling sequence. Another L2CAP QoS signaling message is L2CA_QoSViolationInd, which is used for indicating the address of the remote device that violates a QoS contract.

8.10.4.2 LM QoS signaling

Bluetooth also defines QoS signaling in the LM layer. For example, LMP_quality_of_service and LMP_quality_of_service_req contain the polling interval for the ACL link. These two messages are used for the request and response of the polling interval of the ACL link.

CHAPTER 9

IEEE 802.15

The IEEE 802.15 group focuses on standards that will cover Wireless Personal Area Networks (WPANs). So far IEEE 802.15 has introduced one standard, referred to as IEEE 802.15.1, which standardized parts of Bluetooth. In addition, the group works on additional standards that will include a high data rate WPAN, referred to as IEEE 802.15.3, and a low data rate WPAN, referred to as IEEE 802.15.4. The group is also developing recommended practices to facilitate the coexistence of IEEE 802.15 and IEEE 802.11, referred to as IEEE 802.15.2.

9.1 IEEE 802.15.1

IEEE 802.15.1 standard adopted the Bluetooth Medium Access Control (MAC) and Physical Layer (PHY) specifications in early 2002. It adopted Bluetooth Version 1.1 specifications for lower transport layers (L2CAP, Link Manager Protocol, and Baseband) and radio (see Figure 9.1).

IEEE 802.15.1 PHY and MAC key mechanisms are almost identical to the mechanisms introduced in the Bluetooth specifications. IEEE 802.15.1 specifies the Logical Link Control (LLC) which supports the LLC function. Therefore, the readers can refer to appropriate sections of Chapter 8 for information regarding the physical layer, Baseband, link manager, Logical Link Control and Adaptation Protocol (L2CAP), and Host Control Interface (HCI) in IEEE 802.15.1.

9.2 IEEE 802.15.3

The IEEE 802.15 group is working on a high data rate WPAN that can support connectivity needs between portable devices. This section provides details on a potential future standard. However, the future standard may be completely different from this description.

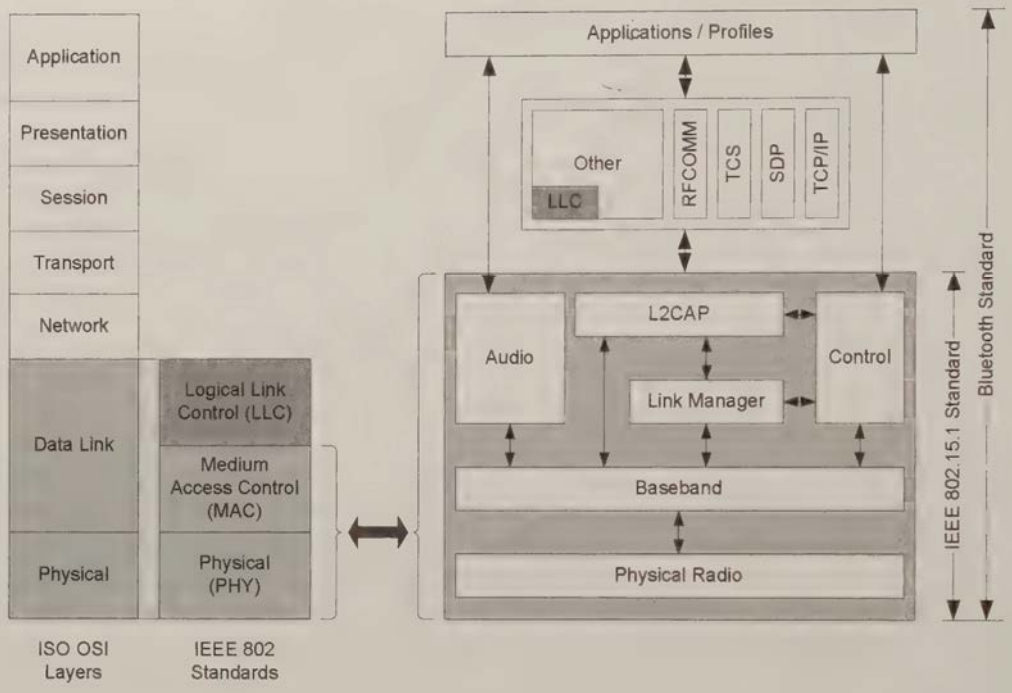


Figure 9.1 Mapping of ISO OSI to IEEE 802.15.1 WPAN Standard

The working group is motivated by the need to distribute video and audio that requires much more bandwidth than IEEEE 802.15.1 can sustain. The proposed radio is based on single carrier Quadrature Amplitude Modulation (QAM) with Trellis Coded Modulation (TCM) working in the 2.4 GHz band. Several selectable speeds are discussed such as 11, 22, 33, 44, and 55 Mbps along with three to four non-overlapping channels. IEEE is also considering another radio technology based on Ultra Wide Band (UWB) that will allow speeds of around 100 Mbps within a 10 meter range and 400 Mbps within a 5 meter range. In addition to data applications, they envision applications such as: high-speed transfer of digital video from a digital camcorder to a TV or PC, interactive games that are media rich, home theater, and delivery of multimedia from a PC to an LCD projector. The MAC provides support for multimedia QoS requirements via a TDMA based superframe architecture with Guaranteed Time Slots (GTSs).

9.2.1 IEEE 802.15.3 Architecture

9.2.1.1 Network Topology

The architecture is based on the concept of a piconet which is established ad hoc and allows a number of independent data devices (DEVs) to establish peer-to-peer communication (see Figure 9.2). It allows mobile devices to join and leave the piconet with short association times. The piconet is confined to a personal area and typically is expected to cover a range of at least 10

meters and possibly up to 70 meters. A piconet is established once a DEV that is capable of becoming the Piconet Coordinator (PNC) begins to transmit control messages called beacons. Through the beacon, the PNC maintains network synchronization by providing timing information. The PNC is also responsible for controlling the channel access, managing the QoS requirements, performing admission control, and assigning time slots for connections between DEVs. Before transmission of data in a piconet, a DEV is required to associate with the PNC and follow the channel access information provided in the beacons.

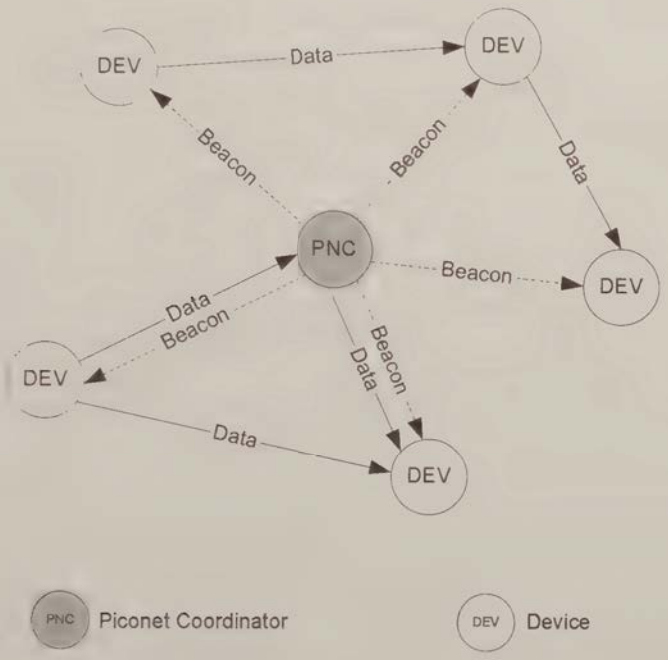


Figure 9.2 IEEE 802.15.3 Piconet Elements

Several piconets can coexist by sharing the same channel frequency. The piconets are combined to cover a larger area and include more DEVs (see Figure 9.3). There is a proposal for a child piconet to be established for an existing piconet. Hence, the child piconet can extend the area of coverage of the existing piconet, increase the number of DEVs to be supported, or shift some tasks to the new piconet. When such a child piconet is established, the existing piconet becomes the parent piconet. The child piconet uses a unique piconet ID (PNID) and functions almost independently. The child PNC handles association, authentication, security, and acknowledgments without the parent PNC. The DEVs that associate to the child piconet can communicate with each other within the child piconet. However, to avoid the interference between the parent piconet and child piconets, the parent PNC has to allocate GTSs or reserve the time slot for the child piconet to access the channel. DEVs of the child piconet are allowed to exchange data only

within the allocated GTS. The child piconet's PNC is also a member of the parent piconet and thus can communicate with the DEVs in both the parent piconet and its own child piconet.

The committee also proposed the idea of establishing a neighbor piconet under an existing piconet that will become the parent piconet. This will provide the mechanism for sharing the frequency spectrum between different piconets when there are no vacant physical radio channels. The neighbor piconet will use a unique PNID. Similar to the child piconet, it will function almost independently from the parent piconet. It will handle association, authentication, security, and acknowledgments without involving the parent piconet. However, the parent piconet has to allocate GTSs for the neighbor piconet to access the channel. Unlike the PNC of the child piconet, the neighbor's PNC is not a member of the parent piconet. The neighbor's PNC is allowed to send to the parent piconet only certain commands such as association request, disassociation, channel time request (CTR), and authentication. The creation process of child and neighbor piconets is described in Section 9.4.2.

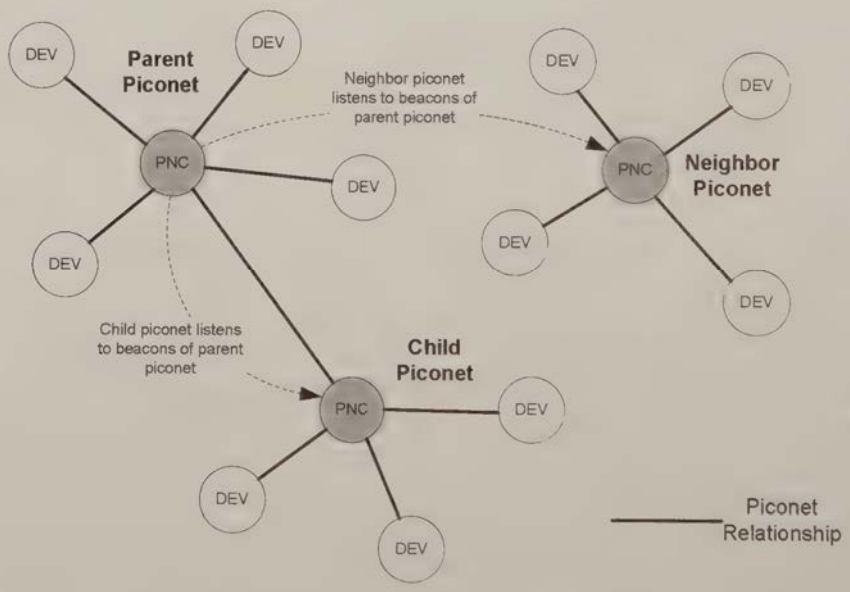


Figure 9.3 Parent/Child/Neighbor Piconets

9.2.1.2 Protocol Stack

As shown in Figure 9.4, the draft standard divided the protocol stack into various layers: frame convergence sublayer (FCSL), MAC sublayer, PHY sublayer and layer-dependent management entity (MLME, PLME). The frame convergence sublayer provides frame multiplexing from different protocols (i.e., 802.2, 1394, USB, etc.). The MAC sublayer defines the following functionalities: association/disassociation, channel access, channel time management, channel synchronization, fragmentation/defragmentation, acknowledgment/retransmission, peer discov-

ery, multirate support, dynamic channel selection, and power management. The PHY sublayer defines the PHY specifications. The Device Management Entity (DME) which is a layer-independent entity provides the status of various layer management entities. The standard does not detail the DME functions, but rather provides a general overview of its functionality. Each entity interacts through a Service Access Point (SAP). Some SAPs (i.e., MAC SAP, PHY SAP, MLME SAP, MLME-PLME SAP) are explicitly defined within the draft standard but some, such as the interface between the MAC and the MLME as well as the interface between PHY and PLME, are not explicitly defined.

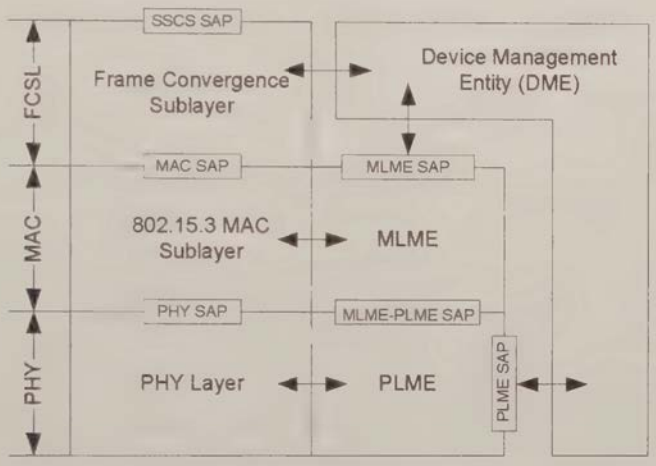


Figure 9.4 IEEE 802.15.3 Protocol Stack

9.3 IEEE 802.15.3 Physical Layer

IEEE 802.15.3 operates in the 2.4 to 2.4835 GHz frequency band. As shown in Table 9.1, there are a total of five channels in two sets of channel assignments: high-density set and 802.11b coexistence set. Each channel has 15 MHz bandwidth. For the high-density set (802.11b does not coexist) four channels are defined while three channels are defined in 802.11b coexistence set. If a DEV detects 802.11b network operating in its area, it should use 802.11b coexistence channel set.

The PNC has the ability to select a channel for the piconet operation. Using information such as interference from other users, other IEEE 802.15.3 piconets, and other unlicensed wireless entities (i.e., IEEE 802.11b) it can make a decision on the best channel for its piconet. The PNC has the capability to dynamically and transparently change the channel on which the piconet is operating without requiring either user intervention or interruption of ongoing applications.