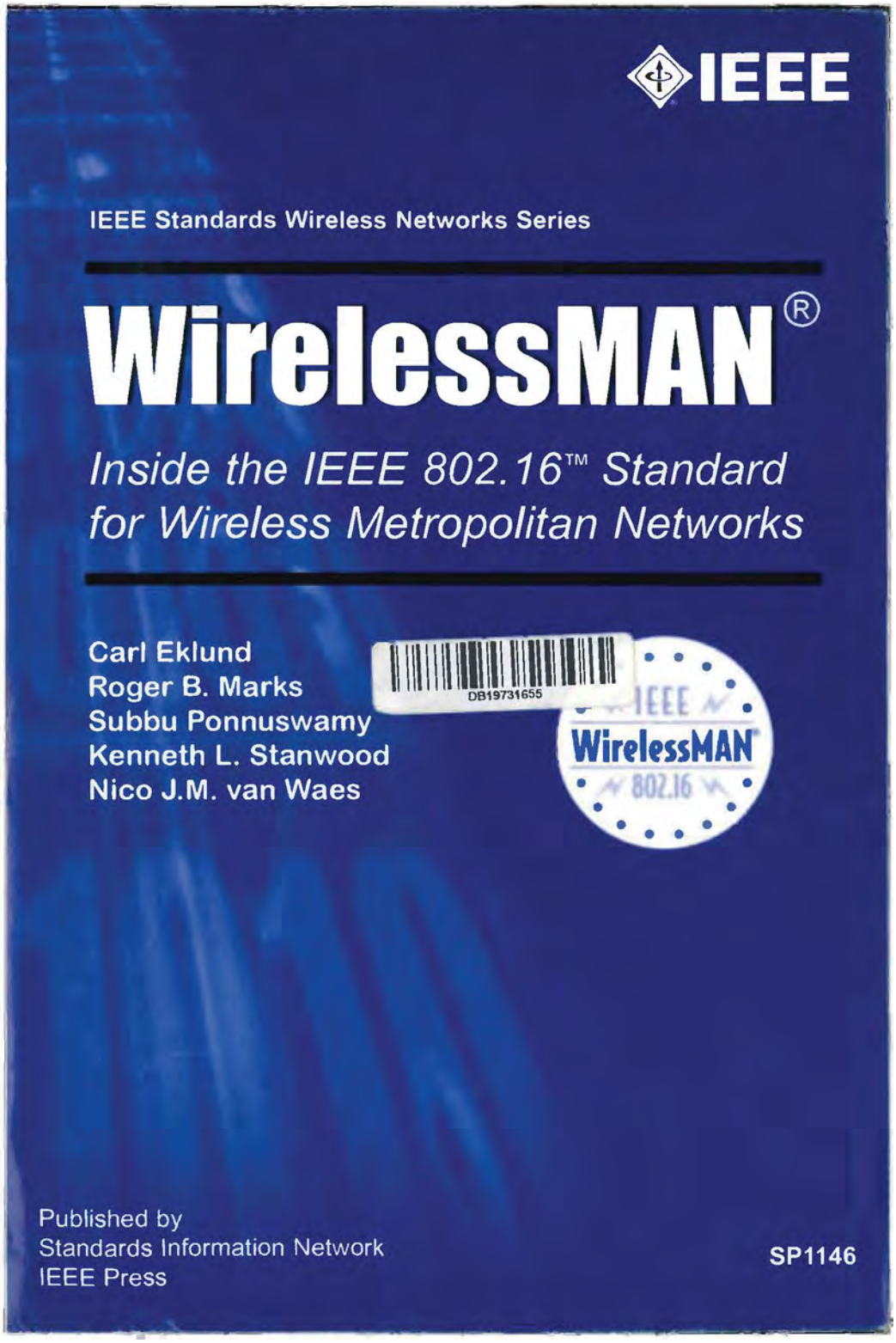

IEEE
IEEE Standards Wireless Networks Series
WirelessMAN®
Inside the IEEE 802.16™ Standard for Wireless Metropolitan Networks
Carl Eklund
Roger B. Marks
Subbu Ponnuswamy
Kenneth L. Stanwood
Nico J.M. van Waes
DB19731655
IEEE
WirelessMAN
802.16
Published by
Standards Information Network
IEEE Press
SP1146

Property of Dat Pham

WirelessMAN®

Inside the IEEE 802.16™ Standard for Wireless Metropolitan Area Networks

Carl Eklund
Roger B. Marks
Subbu Ponnuswamy
Kenneth L. Stanwood
Nico J.M. van Waes



Published by
Standards Information Network
IEEE Press

Trademarks and Disclaimers

IEEE believes the information in this publication is accurate as of its publication date; such information is subject to change without notice. IEEE is not responsible for any inadvertent errors.

Library of Congress Cataloging-in-Publication Data

WirelessMAN : Inside the IEEE 802.16 standard for wireless metropolitan area networks / Carl Eklund ... [et al.].

p. cm. -- (IEEE standards wireless networks series)

Includes bibliographical references and index.

ISBN 0-7381-4842-3

1. Metropolitan area networks (Computer networks). 2. IEEE 802.16 (Standard) 3. Wireless communication systems. I. Eklund, Carl, 1970- II. Series.

TK5105.85.W57 2006
621.384--dc22

2006041723

IEEE
3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2006 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved. Published May 2006. Printed in the United States of America.

No part of this publication may be reproduced in any form, in an electronic retrieval system, or otherwise, without the prior written permission of the publisher.

IEEE, 802, and WirelessMAN are registered trademarks in the U.S. Patent & Trademark Office, owned by the Institute of Electrical and Electronics Engineers, Incorporated (www.ieee.org/).

IEEE Standards designations are trademarks of the IEEE (www.ieee.org/).

DOCSIS is a registered trademark of Cable Labs (http://www.cablelabs.org).

Wi-Fi is a registered trademark of the Wi-Fi Alliance (http://www.wi-fi.org/).

WiMAX Forum is a registered trademark of the WiMAX Forum (http://www.wimaxforum.org).

The following figures in this book are the property of Doceotech, Incorporated, and are reprinted with permission: 1-1, 3-5, 6-3 through 6-7, 6-9 through 6-11, 6-21, 6-22, and 9-4.

Jennifer McClain Longman, Managing Editor
Linda Sibilia, Cover Designer

Review Policy

Standards Information Network/IEEE Press publications are not consensus documents. Information contained in this and other works has been obtained from sources believed to be reliable and reviewed by credible members of IEEE Technical Societies, Standards Committees, and/or Working Groups, and/or relevant technical organizations. Neither the IEEE nor its authors guarantee the accuracy or completeness of any information published herein, and neither the IEEE nor its authors shall be responsible for any errors, omissions, or damages arising out of the use of this information.

Likewise, while the authors and publisher believe that the information and guidance given in this work serve as an enhancement to users, all parties must rely upon their own skill and judgment when making use of this information and guidance. Neither the authors nor the publisher assumes any liability to anyone for any loss or damage caused by any error or omission in the work, whether such error or omission is the result of negligence or any other cause. Any and all such liability is disclaimed.

This work is published with the understanding that the IEEE and its authors are supplying information through this publication, not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought. The IEEE is not responsible for the statements and opinions advanced in the publication.

The information contained in Standards Information Network/IEEE Press publications is reviewed and evaluated by peer reviewers of relevant IEEE Technical Societies, Standards Committees and/or Working Groups, and/or relevant technical organizations. The authors addressed all of the reviewers' comments to the satisfaction of both the Standards Information Network/IEEE Press and those who served as peer reviewers for this document.

The quality of the presentation of information contained in this publication reflects not only the obvious efforts of the authors, but also the work of these peer reviewers. The IEEE Press acknowledges with appreciation their dedication and contribution of time and effort on behalf of the IEEE.

To order IEEE Press Publications, call 1-800-678-IEEE.

Print: ISBN 0-7381-4842-3 SP1146

See other IEEE standards and standards-related product listings at http://standards.ieee.org/

Dedication

The authors dedicate this effort to the many members of, and participants in, the IEEE 802.16 Working Group on Broadband Wireless Access. Their tireless efforts since 1998 have led to the completion of a dozen standards with the potential to bring the world closer together. At the very least, the project has already brought a world of participants closer together.

Acknowledgments

The authors acknowledge the detailed reviews provided by Dr. David R. Smith and Dr. Claude Weil of the (U.S.) National Institute of Standards and Technology (NIST), as well as the assistance of several anonymous reviewers.

Portions of this work were contributed by NIST and are not subject to copyright protection in the United States.

The authors further acknowledge Doceotech, Incorporated, which provided some of the figures in this book. These figures, reprinted with permission, are drawn from one of the company's courses providing training for engineers on wireless technologies.

0004

Authors



Carl Eklund received his M.S. in engineering physics from Helsinki University of Technology in 1996. He joined the Communication Systems Laboratory of Nokia Research Center in 1998, working mainly on radio protocol design and standardization. In the IEEE 802.16 effort, he chaired the MAC Task Group that developed the IEEE 802.16 medium access control layer (MAC) protocol for IEEE Std 802.16-2001. He also served as the technical editor for the protocol implementation conformance statement (PICS) and test suite structure and test purposes (TSS&TP) specifications for IEEE Std 802.16-2001. Eklund currently is a principal engineer in the Radio Communications Laboratory of Nokia Research Center, Helsinki, Finland. Since October 2005, he has been heading the research and standardization program for WiMAX and IEEE 802.16 in Nokia.



Roger B. Marks initiated, in 1998, the effort leading to the formation of the IEEE 802.16 Working Group on Broadband Wireless Access, chairing it since inception and serving as Technical Editor of the group's first two standards. He also serves actively on the IEEE 802 Executive Committee and holds the position of China Liaison Official. Marks is a physicist with the (U.S.) National Institute of Standards and Technology (NIST) in Boulder, Colorado, USA. He received his A.B. in physics in 1980 from Princeton University and his Ph.D. in applied physics in 1988 from Yale University. A Fellow of the IEEE and an IEEE Distinguished Lecturer, Marks developed the IEEE Radio and Wireless Conference and chaired it from 1996 through 1999. He is the author of over 90 publications and the recipient of numerous awards, including the Individual Governmental Vision Award from the Wireless Communications Association and the IEEE Technical Field Award in measurement technology. He has received the U.S. Department of Commerce Gold, Silver (three times), and Bronze Medals.



Subbu Ponnuswamy was one of the early participants in the IEEE 802.16 Working Group and a contributor to the IEEE 802.16 and IEEE 802.11 standards. He is also a coauthor of a WiMAX course for development engineers, offered by Doceotech. He has many years of industry experience in the design and development of wireless local area network (LAN) and metropolitan area network (MAN) products, including those based on the IEEE 802.16 and IEEE 802.11 standards. As the director of engineering at Kiwi Networks, Ponnuswamy led the design and development of interference-resilient IEEE 802.16 and IEEE 802.11 systems in the license-exempt bands for indoor and outdoor applications. He also led IEEE 802.11 MAC application-specific integrated circuit (ASIC) and software development at Vivato for smart antenna systems. During his tenure at Malibu Networks, he designed and developed a quality-of-service-centric broadband wireless MAC. He has also held various technical positions with Honeywell, Sequent Computer Systems, and Lincom Wireless. He is currently with Aruba Networks. Ponnuswamy is the author of many publications and patents in the areas of wireless communication, real-time systems, and multiprocessor communication networks. He graduated with an M.S. in computer engineering from Wayne State University and a B.E. in electronics and communication engineering from the University of Madras, India. He is a member of the Association for Computing Machinery (ACM) and the IEEE Communications Society.



Kenneth L. Stanwood is president and chief executive officer of Cygnus Communications, which makes products for wireless multimedia distribution. He was previously chief technology officer of Ensemble Communications, which produced local multipoint distribution services (LMDS) equipment and provided key technology to IEEE 802.16 and WiMAX. As a representative of Ensemble, Stanwood was one of the founders of the WiMAX Forum® and served on its board of directors. Stanwood is vice-chair of the IEEE 802.16 Working Group and has been involved with IEEE 802.16 and the European Telecommunications Standards Institute (ETSI) Broadband Radio Access Networks (BRAN) Technical Committee for over 6 years. He was a primary designer of the

IEEE 802.16 MAC. He holds 11 patents and has numerous patent applications, all related to broadband wireless access. He received his master's degree from Stanford University.



Nico J.M. van Waes received an M.S.E.E from the Technical University Delft in the Netherlands in 1994 and a Ph.D. from New Jersey Institute of Technology in 1998. He joined the Wireless Router Division of Nokia Networks in 1999 as a systems engineer, working primarily on physical layer (PHY) and radio frequency (RF) issues as well as standardization. From 1999 till 2004, van Waes held various standards-related public positions such as chief technical editor of IEEE Std 802.16a™, IEEE P802.16.2a, and early versions of IEEE P802.16d; area coordinator and editor for ETSI BRAN HiperMAN; and chair of the OFDM Forum's fixed wireless access (FWA) working group. From 2004 till 2005, he led Nokia Research Center's efforts in IEEE P802.11n standardization. Since early 2006, van Waes has been a manager with Nokia IPR, responsible among others for the IEEE 802.16 and IEEE 802.11 portfolios. He has half a dozen patents filed and is the author of several published papers.

Foreword

Because of the explosive global growth of the Internet and its applications, this authoritative, comprehensive, and well-written insider's guide to broadband wireless standards is truly a book whose time has come. Published by the Standards Information Network/IEEE Press, a unit of the Institute of Electrical and Electronics Engineers (IEEE), *WirelessMAN* serves as a companion to IEEE Std 802.16, the WirelessMAN standard for broadband wireless metropolitan area networks. This book provides an overview of the technology, explaining the rationale behind the choices made. This overview of the complex technical specification is valuable for the tens of thousands of engineers involved in R&D and deployment of relevant technologies—as well as to less technical opinion leaders in related fields of marketing, sales, finance, government affairs, and media/public relations. *WirelessMAN* is so clearly organized and written that it can be a rewarding reading experience even for those with no formal technical training.

The authors are and have been leaders of the related standardization projects. IEEE Std 802.16 was developed by the IEEE 802.16 Working Group on Broadband Wireless Access, beginning in 1999. The book's authors include the IEEE 802.16 founding chair, Dr. Roger B. Marks; its vice chair, Kenneth L. Stanwood; and three of its most visionary and dedicated practitioners, Carl Eklund, Subbu Ponnuswamy, and Nico J.M. van Waes. Those involved in standards work can appreciate that it is a process whose success depends upon far more than just vision and technical expertise. That's only the beginning. To be as successful as the WirelessMAN standards effort has been, it's necessary also to draw on the skill sets commonly associated with master politicians and corporate strategists because the stakes can be so huge—especially in this instance, where the goal is nothing short of gradually replacing multibillion-dollar proprietary technologies with lower cost, standards-based interoperable products. If the upside is high for some players, so is the downside for others. In other words, creating a successful standard is

not in everyone's interest, and so whatever method of collegial input is used may also be abused.

The IEEE 802.16 Working Group has made impressive progress in its standards effort, as measured by objective data. It has held more than 43 weeklong sessions around the world, from the group's official founding in 1999 to the publication of this book in 2006. By working cooperatively, the group has developed and advanced its standards in near-record time. The current version of its fixed wireless standard was most recently updated as IEEE Std 802.16-2004 [B20], and an amendment adding mobility support (IEEE Std 802.16e™-2005 [B24]) was published in late 2005. The IEEE 802.16 standard is at the core of the fast-growing certification effort by the WiMAX Forum®, whose expanding activities have been generating news and magazine articles at the rate of six hundred a month at times.

By attending the organizational strategy meeting in July 1998 of the predecessor group that led to the IEEE 802.16 Working Group, I had the privilege of observing the growth of this process right from the beginning despite my lack of any formal training in the relevant technology. Fortunately, I did have good advice from my mentor in this area, Dr. Weston E. Vivian, the longtime University of Michigan professor who holds the distinction of being the first person ever elected to the U.S. Congress with a doctorate in either engineering or science. His counsel was to "sit in on every technical meeting you can – and don't say much." The assumption was that over time, even I would learn something. That strategy meeting was called by Roger Marks at the Boulder, Colorado, offices of the National Institute of Standards and Technology (NIST) to create what became known as the National Wireless Electronic Systems Testbed before its assimilation the next year into the IEEE standards process. The few others at the meeting had enormous subject-matter expertise, most with doctorates in electronic engineering and affiliations with major companies. Despite the additional handicap of not being in a position to make substantive commitments because Wireless Communications Association International (WCA) is technology neutral, I attended as an observer, encouraged by knowledge that standards were regarded by our members as potentially vital, but also extremely difficult to achieve in practice.

We at WCA have been extremely impressed by the dedication of the WirelessMAN pioneers, including its founding chair. Although WCA continues to adhere to a technology-neutral standpoint in order to showcase all relevant technologies on a fair basis, it seems appropriate nevertheless to share a few personal observations at this juncture: First, the leadership of this group is particularly dedicated, well-organized, and enthusiastic about creating new networks. These standards meetings are all-day, weeklong affairs, and more than once I have called upon leaders in the middle of their night to learn of developments. Second, and more important, *WirelessMAN*, an authoritative insider's account, is certain to prove valuable as a guide to understanding the history and implications of the technology—not just the IEEE 802.16 version, but in the context of competing and complementary broadband technologies, both wireless and wireline. The book is highly recommended for designers of components, systems, and test equipment based on the WirelessMAN standards, plus a variety of other telecommunications professionals and students. This is an important and useful book for anyone in this fast-exploding field.

Andrew Kreig
President and Chief Executive Officer
Wireless Communications Association International
1333 H Street, NW; Suite 700 West
Washington, DC 20005
United States of America

Preface

Broadband access has become a critical bottleneck in the ongoing information technology revolution across the globe. The growing popularity of multimedia applications has made the requirements on broadband access even more demanding. In order to serve this growing market need, the reach of existing broadband networks needs to be extended, and new broadband infrastructures need to be deployed in developing nations where no wired broadband infrastructures exist.

The advent of standards-based and interoperable wireless local area networks (WLANs) and wireless personal area networks (WPANs) has resulted in widespread adoption of wireless networks in homes as well as in consumer and enterprise markets. The productivity gains and increased flexibility that comes with wireless have transformed the way many businesses operate.

Broadband wireless access (BWA) is the next logical step in extending fiber optic backbone networks to enhance broadband proliferation and to provide broadband access in areas where no other methods are available. BWA is also attractive as a network infrastructure that can be deployed much more quickly than its wired counterparts. Moreover, broadband wireless has the potential to offer capacity on demand.

Proprietary BWA solutions have existed for some time, with limited market penetration in licensed and license-exempt bands. However, the lack of a global standard and lack of multivendor interoperable equipment have hampered widespread adoption. Given the infrastructure-intensive nature of broadband wireless deployments, widespread adoption requires interest from a wide range of service providers. All of the potential providers, large or small, are seeking standardized rather than proprietary solutions. With nomadic and mobile systems, roaming operations force questions of cross-vendor interoperability to the forefront. However, even for non-nomadic fixed broadband wireless networks, in which roaming is virtually nonexistent, operators are hesitant to accept unnecessary risk in the current economic

climate, and the use of proprietary equipment from a small number of vendors is an unnecessary risk. Good standards activities, such as those that are well established in IEEE 802®, not only bring operators reasonably priced equipment but also put them on a curve of declining costs and growing performance. The IEEE 802.16 standard for wireless metropolitan area networks (MANs) has been developed by the IEEE 802.16 Working Group to address this BWA market. The activity has been in progress since 1999; it has been highly productive, and the standard continues to evolve.

As you will find in this book, IEEE 802.16 is quite different from other IEEE 802 standards. The technology is more oriented toward carrier-class services whose infrastructure development is designed and deployed by professionals affiliated with experienced telecommunications operators. Also, compared to some other IEEE 802 standards, IEEE 802.16 leaves more of the implementation decisions to the implementor, particularly at the base station (BS) side. This provides the system developer with the opportunity to develop a simple, low-cost system or to apply extra effort to create a substantially better performing system that may fetch a higher price. The goal is a wide deployment of a broad range of IEEE 802.16-based systems that are interoperable but highly differentiated. This should help to keep the market dynamic and innovative and provide many attractive options to the carriers.

Objective

The objective of this book is to provide a detailed overview of the IEEE 802.16 standard for fixed BWA, including detailed descriptions of the IEEE 802.16 medium access control layer (MAC) and physical layer (PHY) operation, with emphasis on the design and the technology behind the standard. This book explains why certain design choices were made and how recent technological developments, real-world experience, and lessons learned from previous projects were used throughout the IEEE 802.16 development process to make critical tradeoffs.

The book also reviews the chronological development of the IEEE 802.16 MAC and PHY and the rationale behind the development decisions. It provides a summary of ongoing projects, related standards, and future extensions to IEEE 802.16. The optional and mandatory parts of the

IEEE 802.16 standard are clearly identified and explained throughout the book, wherever applicable.

Organization

Chapter 1 provides an introduction to BWA and identifies the key market segments and basic requirements. Chapter 2 overviews the IEEE 802.16 Working Group, including the organizations under which it works and the procedures it uses, summarizing the history of the group and its past and current projects. Chapter 3 defines the basic MAC, PHY, radio frequency (RF) and other protocol concepts necessary for understanding the IEEE 802.16 standard. The final introductory chapter, Chapter 4, describes the overall architecture of IEEE 802.16 and introduces various components and key features of IEEE 802.16 in detail.

The rest of the book describes the components of the IEEE 802.16 architecture in a top-down fashion. Chapter 5 describes the asynchronous transfer mode (ATM) convergence sublayer (CS) and packet convergence sublayer (PCS) of IEEE Std 802.16. In Chapter 6, the basic concepts of the IEEE 802.16 MAC, supporting all of the IEEE 802.16 PHY alternatives, are covered. Chapter 7 provides details of the IEEE 802.16 MAC operation, including network entry, initialization, PHY support, automatic repeat request (ARQ) and quality of service (QoS). The security sublayer, including encryption methods and key exchange mechanisms, is described in Chapter 8. Chapter 9 describes the mesh extensions to the IEEE 802.16 MAC and the additional scheduling methods defined to support the mesh option.

The next three chapters describe three major PHY alternatives specified in IEEE 802.16. Chapter 10 addresses the single-carrier (WirelessMAN-SC) PHY specified for use above 10 GHz. Chapter 11 describes the orthogonal frequency division multiplexing (WirelessMAN-OFDM) PHY for frequencies below 11 GHz, and Chapter 12 describes the orthogonal frequency division multiple access (WirelessMAN-OFDMA) PHY for the same frequencies.

Chapter 13 explores the IEEE 802.16 standard's support for multiple antenna systems, including adaptive antenna systems (AAS), as an extension of the IEEE 802.16 point-to-multipoint (PMP) topology. Chapter 14 addresses the

performance of the MAC and various PHYs. A summary of the IEEE 802.16 conformance standards and an introduction to the interoperability work being done by the WiMAX Forum are given in Chapter 15. Chapter 16 provides a summary of related standards and standardization activities.

Appendix A provides a list of IEEE 802.16 headers, subheaders, and MAC management messages for ready reference.

Reading and style

This book is intended for anyone who wishes to understand the IEEE 802.16 standard and operation, including designers, engineers, students, and deployment professionals. This book is self-contained, and no specific knowledge of any other wireless protocols is assumed. However, a basic knowledge of communication protocols and concepts is necessary for understanding the key MAC and PHY chapters of this book. An extensive bibliography is given at the end for anyone who would like to read more on specific topics that are not covered here in detail.

While the introductory chapters provide basic information for all types of readers, the MAC- and PHY-specific chapters can be treated independently, with the exception of the MAC-PHY interface, by designers who are interested in designing only the MAC or PHY portion of a IEEE 802.16 system. The mesh and AAS chapters require the understanding of the MAC architecture and PHY interface defined in earlier chapters. This book can be used for self-study of the IEEE 802.16 standard, as a reference, or as a designer's handbook.

Advanced readers should look at the book as a companion to IEEE Std 802.16. Readers should note that IEEE 802 standards are available for free download through the Get IEEE 802® program <http://standards.ieee.org/getieee802>. Readers may also wish to consult the web pages of the IEEE 802.16 Working Group on Broadband Wireless Access <http://WirelessMAN.org>, which include thousands of working group and contributed documents upon which the standard, including its amendments, are based.

Table of Contents

List of Figures xxvii

List of Tables xxxi

Acronyms and abbreviations xxxiii

Chapter 1 Broadband wireless access (BWA):
Applicable market segments and requirements 1
Commercial fixed broadband wireless: fiber extension 6
Residential fixed broadband wireless: digital subscriber line (DSL) and cable modem alternative 8
Quality of service (QoS) 11
Throughput requirements 11

Chapter 2 IEEE 802.16 standards:
The working group and documents 13
Background 13
IEEE Standards Association (IEEE-SA) 13
IEEE 802® LAN/MAN Standards Committee (LMSC) 15
Standards development in IEEE 802 16
Study group stage 16
Working group development of draft 16
IEEE 802.16 Working Group: Overview 19
IEEE 802.16 Working Group: History 19
Technical progress in IEEE 802.16 Working Group 21
Air interface: IEEE Std 802.16 23
Conformance: IEEE 802.16/Conformance0X 26
Coexistence: IEEE Std 802.16.2™ 27

Chapter 3 Basic concepts and definitions:
Wireless protocol and communication concepts 29
Frequency bands 30
Channels 30
Licensed and license-exempt spectrum 30
Spectrum and standardization 32
Coexistence 32
Types of wireless networks 33

Fixed and mobile networks . . . 33
Nomadic and portable networks . . . 34
Wireless network topologies . . . 34
RF propagation . . . 36
LOS and NLOS . . . 36
Multipath . . . 38
Fading . . . 39
Antennas . . . 40
Antenna parameters . . . 41
Directional and sectorized antennas . . . 42
Diversity . . . 42
Temporal diversity . . . 43
Frequency diversity . . . 43
Spatial diversity . . . 43
Polarization diversity . . . 44
Angle diversity . . . 45
Multiple antenna systems . . . 45
Adaptive antenna systems (AAS) . . . 45
Multiple-input, multiple-output (MIMO) . . . 46
Impact of antenna technologies on protocol design . . . 47
Antenna design for fixed and mobile devices . . . 47
Physical layer (PHY) . . . 48
Forward error correction (FEC) . . . 48
Single carrier and multicarrier . . . 49
Duplexing, multiplexing, and multiple access . . . 49
Duplexing . . . 50
Multiplexing . . . 52
Centralized and distributed multiple access schemes . . . 52
Time division multiple access (TDMA) and frequency division multiple access (FDMA) . . . 53
Orthogonal frequency division multiple access (OFDMA) . . . 53
Code division multiple access (CDMA) . . . 55
Data units . . . 56
Quality of service (QoS) . . . 57
Per-flow and per-class QoS . . . 60
Wireless QoS . . . 61
Medium access control layer (MAC) . . . 62
Fragmentation and packing . . . 62
Automatic repeat request (ARQ) . . . 64

Chapter 4 IEEE 802.16 architecture:

Overview and key features . . . 67
Reference model . . . 68
Base station (BS) and subscriber station (SS) . . . 71
Convergence sublayer (CS) architecture . . . 73
Framing and duplexing . . . 74
Physical slots (PS) and mini-slots . . . 76
TDD framing . . . 76
FDD framing . . . 78
Time relevance . . . 81
Subscriber-level adaptive PHY . . . 84
Fixed TDD vs ATDD . . . 84
Framed PHY . . . 85
MAC efficiency . . . 85
Mesh . . . 87
Directed mesh . . . 88
Quality of service (QoS) . . . 88
Security sublayer . . . 89
Automatic repeat request (ARQ) . . . 90
Hybrid automatic repeat request (HARQ) . . . 92
Physical layer (PHY) . . . 93
Multipath mitigation . . . 94
Mandatory and optional components . . . 94
Bit ordering . . . 98

Chapter 5 Convergence sublayers (CSs):

Support for multiple protocol transport . . . 99
ATM CS . . . 99
Packet convergence sublayer (PCS) . . . 101
Classification . . . 102
Payload header suppression (PHS) . . . 104

Chapter 6 MAC basics:

Concepts, connections, formats, and headers . . . 107
Connections and addressing . . . 107
Service flows and service flow identifiers (SFIDs) . . . 109
CID allocation . . . 109
MAC headers and subheaders . . . 113
Stand-alone MAC headers . . . 113
BW request header . . . 114
MPDU header . . . 115

Generic MAC header . . . 115
MAC header demultiplexing . . . 118
MAC subheaders . . . 119
Fragmentation subheader (FSH) . . . 120
Packing subheader (PSH) . . . 121
Grant management subheader (GMSH) . . . 122
Mesh subheader (MSH) . . . 123
Fast-feedback allocation subheader (FFSH) . . . 124
ARQ feedback . . . 125
Data and management PDU construction . . . 125
Simple MPDU . . . 125
Subheader ordering . . . 126
ARQ blocks . . . 127
Fragmentation . . . 129
Fragmentation without ARQ . . . 130
Fragmentation with ARQ . . . 130
Packing . . . 131
Packing of fixed-length SDUs . . . 131
Packing of variable-length SDUs . . . 132
Packing with fragmentation . . . 132
Packing and ARQ . . . 134
Packing of ARQ payload . . . 134
Concatenation . . . 135
MPDU encryption and CRC . . . 135
MAC management . . . 136
ARQ . . . 137
ARQ block-based retransmissions . . . 137
ARQ Feedback information element (IE) . . . 139
ARQ feedback payload . . . 143
Hybrid automatic repeat request (HARQ) . . . 143
Compact MAP IE . . . 144
HARQ Control IE . . . 144
Construction of HARQ packets . . . 145
Reduced connection identifier (RCID) . . . 145
HARQ acknowledgments . . . 147

Chapter 7 MAC operation:
Radio control, QoS, and ARQ . . . 149
Network entry and initialization . . . 149
Scanning and synchronization to the DL . . . 149
Initial ranging . . . 149

SS basic capability negotiation . . . 150
Authorization, security association (SA) establishment, and key exchange . . . 152
Registration . . . 153
Establish IP connectivity . . . 153
Dynamic service establishment . . . 154
PHY maintenance . . . 155
MAPs and channel descriptors . . . 156
Periodic ranging . . . 163
Burst profile changes . . . 167
QoS and service flows . . . 171
Dynamic service establishment and deletion . . . 171
QoS model . . . 172
QoS and traffic parameters . . . 173
SFID and CID . . . 173
Service class name . . . 173
QoS parameter set type . . . 174
Traffic priority . . . 176
Maximum sustained traffic rate . . . 176
Maximum traffic burst . . . 177
Minimum reserved traffic rate . . . 177
Minimum tolerable traffic rate . . . 177
Vendor-specific QoS parameters . . . 178
Service flow scheduling type . . . 178
Request/transmission policy . . . 179
Tolerated jitter . . . 179
Maximum latency . . . 179
Fixed-length vs variable-length SDU indicator . . . 180
SDU size . . . 180
Interactions between QoS, CAC, and adaptive PHY . . . 181
Determining available bandwidth . . . 181
Bandwidth on demand setting the basis for CAC . . . 182
Adaptive CAC philosophies for adverse conditions . . . 187
Multicast connections . . . 189
BW request/grant . . . 190
Scheduling . . . 193
Unicast polling . . . 197
Broadcast polling . . . 199
Multicast polling groups . . . 200
Clock comparison . . . 201
ARQ operation . . . 203
ARQ parameters and timers . . . 203

ARQ protocol messages ... 206
ARQ Feedback management message ... 206
ARQ Discard management message ... 206
ARQ Reset management message ... 206
BSN comparison ... 207
ARQ transmitter ... 207
ARQ feedback processing ... 210
ARQ receiver ... 210
ARQ feedback generation ... 213
ARQ state machine reset and resynchronization ... 213
Interaction with scheduler ... 215
HARQ operation ... 216
HARQ parameters ... 216
Subpacket transmission and acknowledgment generation ... 216
Performance and QoS implications ... 217

Chapter 8 Security:
PKM protocol and cryptographic methods ... 219
Security associations (SAs) and cryptographic suites ... 219
Encrypted MPDUs ... 220
Data encryption with DES in CBC mode ... 220
AES-CCM ... 221
Key management ... 224
AK management ... 224
TEK management ... 226

Chapter 9 Mesh:
MAC and PHY extensions for mesh ... 229
Introduction ... 229
Logical mesh ... 232
Mesh connections and addressing ... 234
Network configuration ... 234
Network entry ... 236
Scanning for active networks and coarse synchronization ... 237
Obtaining network parameters ... 237
Opening sponsor channel ... 238
Distributed scheduling ... 239
Centralized scheduling ... 241
Directed mesh and point-to-point (PtP) ... 244

Chapter 10 PHY: WirelessMAN-SC:
Single-carrier PHY for 10–66 GHz ... 247
Frame structure ... 248
DL channel encoding ... 252
UL channel encoding ... 256
Control mechanisms ... 259
WirelessMAN-SCa ... 260

Chapter 11 PHY: WirelessMAN-OFDM:
Multicarrier PHY for frequencies below 11 GHz ... 261
Waveform construction ... 261
Selection of the OFDM waveform ... 261
Selection of FFT size ... 262
The OFDM waveform ... 265
Subchannelization ... 266
Preambles ... 270
Frame structure ... 272
Point-to-multipoint (PMP) ... 272
Mesh ... 275
Channel encoding ... 277
Randomization ... 277
Forward error correction (FEC) ... 278
Reed-Solomon concatenated with convolutional coding (RS-CC) ... 278
Block turbo codes (BTCs) ... 279
Convolutional turbo codes (CTCs) ... 280
Interleaving ... 281
Modulation ... 281
Control mechanisms ... 282
Ranging ... 282
BW requests ... 284
Power control ... 285

Chapter 12 PHY: WirelessMAN-OFDMA:
Multicarrier PHY for frequencies below 11 GHz ... 287
Introduction ... 288
Selection of OFDMA waveform ... 288
Time-frequency mapping ... 290
Permutation examples ... 294
FUSC ... 294
PUSC, DL ... 296
PUSC, UL ... 298

Frame structure ... 300
Point-to-multipoint (PMP) ... 300
Preambles ... 301
Channel encoding ... 302
Randomization ... 302
Forward error correction (FEC) ... 302
Convolutional coding (CC) ... 303
BTCs and CTCs ... 303
LDPCC ... 303
Interleaving ... 305
Modulation ... 305
Control mechanisms ... 306
Fast-feedback ... 306
CDMA ranging and BW requests ... 306
PAPR reduction/safety zone ... 308
HARQ support ... 309

Chapter 13 Multiple antenna systems:
Support for advanced antennas ... 311
Adaptive antenna systems (AAS) ... 311
AAS support in IEEE Std 802.16 ... 313
DL and UL framing ... 314
MAC service and control functions ... 316
AAS MAC management messages ... 316
Channel state information ... 317
AAS DL synchronization and initial ranging ... 317
AAS BW requests ... 318
AAS support in OFDM PHY ... 319
AAS support in OFDMA PHY ... 320
Open-loop transmit diversity ... 321
STC support in OFDM PHY ... 322
STC, FHDC, and MIMO support in OFDMA PHY ... 323
Closed-loop transmit diversity ... 324
Precoding ... 325
Antenna selection ... 326
Antenna grouping ... 326

Chapter 14 Performance analysis:
MAC and PHY performance and throughput ... 327
Introduction ... 327
WirelessMAN-OFDM, fixed operation ... 327

Capacity analysis ... 327
MAC performance ... 329
Full-duplex with single burst profile ... 331
Mixed full-duplex/half-duplex with burst profiles case ... 335
WirelessMAN-OFDM, mobile operation ... 338
Basic PHY performance ... 338
Capacity analysis ... 340
WirelessMAN-OFDMA, mobile operation ... 343
Overhead and capacity ... 343
WirelessMAN-OFDMA vs high-speed DL packet access (HSDPA) ... 348
Basic WirelessMAN-OFDMA PHY performance ... 349
MIMO capacity ... 355

Chapter 15 Conformance and interoperability:
Conformance standards and testing ... 357
The WiMAX Forum ... 357
Conformance test standards for WirelessMAN-SC ... 360
The WiMAX Forum's move to lower frequencies ... 360
Lower frequency profiles and test specifications ... 361
WiMAX Forum conformance testing for fixed access ... 362
WiMAX Forum and mobile broadband wireless access ... 363

Chapter 16 Related standards:
Other wireless standards with similar applications ... 365
IEEE Std 802.11 ... 365
IEEE 802.11 MAC ... 366
Medium access control layer (MAC) ... 366
MAC overhead ... 367
MAC summary ... 368
IEEE 802.11 PHY ... 369
IEEE 802.11 extensions ... 370
Using IEEE Std 802.11 as a MAN ... 371
Quantitative comparison of IEEE Std 802.11 and IEEE Std 802.16 ... 371
IEEE 802.20 Working Group ... 373
IEEE 802.22 Working Group ... 374
ETSI BRAN ... 375
ETSI BRAN HiperACCESS ... 376
ETSI BRAN HiperMAN ... 377
Other regional standards activities ... 378
Korean Telecommunication Technology Association (TTA) and WiBro ... 378

China Communications Standards Association (CCSA) 379
International Telecommunications Union (ITU) 379

Appendix A IEEE 802.16 headers, subheaders, and management messages 383

Bibliography 387

Index 393

List of Figures

Figure 1–1: Typical IEEE 802.16 deployment scenarios 7
Figure 2–1: IEEE 802.16 Working Group attendance 20
Figure 2–2: IEEE 802.16 project timeline 22
Figure 3–1: Fresnel zone 37
Figure 3–2: TDD illustration 50
Figure 3–3: Full-duplex FDD illustration 51
Figure 3–4: H-FDD illustration 51
Figure 3–5: OFDM and OFDMA 54
Figure 3–6: PDUs, SDUs and SAPs 57
Figure 4–1: IEEE 802.16 reference model 69
Figure 4–2: IEEE 802.16 framing 74
Figure 4–3: PMP TDD frame structure 77
Figure 4–4: PMP FDD frame structure 79
Figure 4–5: H-FDD framing 80
Figure 4–6: Full-duplex FDD framing 80
Figure 4–7: Maximum time relevance of UL and DL MAPs for TDD 81
Figure 4–8: Minimum time relevance of UL and DL MAPs for TDD 82
Figure 4–9: Maximum time relevance of DL and UL MAPs for FDD 83
Figure 4–10: Minimum time relevance of DL and UL MAPs for FDD 83
Figure 5–1: ATM CS PDU format 100
Figure 5–2: CS PDU format for VP-switched ATM connections 100
Figure 5–3: CS PDU format for VC-switched ATM connections 100
Figure 5–4: MSDU format for PCS 101
Figure 5–5: Classification and CID mapping (SS to BS) 102
Figure 5–6: PHS operation, with masking 105
Figure 6–1: BW request header 114
Figure 6–2: Generic MAC header 116
Figure 6–3: FSHs 120
Figure 6–4: PSHs 122
Figure 6–5: GMSHs 123
Figure 6–6: MSH 124
Figure 6–7: FFSH 124
Figure 6–8: MPDU structure 125
Figure 6–9: MPDU with subheaders 126
Figure 6–10: Ordering of MAC subheaders 128
Figure 6–11: ARQ blocks 129

Figure 6–12: MPDU with fragmentation 130
Figure 6–13: Packing of fixed-length SDUs 131
Figure 6–14: Packing of variable-length SDUs 132
Figure 6–15: Packed MPDU with two fragments 133
Figure 6–16: Packed MPDU with one fragment 133
Figure 6–17: ARQ feedback and packing 134
Figure 6–18: Concatenation of MPDUs 135
Figure 6–19: Format of the MAC management message 136
Figure 6–20: ARQ blocks and retransmissions 138
Figure 6–21: ARQ Feedback IE 140
Figure 6–22: Block sequence acknowledgment MAP formats 143
Figure 6–23: Construction of HARQ packet 145
Figure 6–24: RCID decoding 146
Figure 7–1: SS basic capability negotiation 151
Figure 7–2: SS registration 154
Figure 7–3: DSA message flow (BS-initiated) 155
Figure 7–4: TDD frame format 157
Figure 7–5: FDD frame format 158
Figure 7–6: FDD DL subframe structure 159
Figure 7–7: Practical TDD UL minimum-maximum relevance 161
Figure 7–8: FDD logical offset 162
Figure 7–9: Periodic ranging opportunity allocation at BS 164
Figure 7–10: Periodic ranging receiver processing at BS 165
Figure 7–11: Periodic ranging at SS 166
Figure 7–12: CDMA periodic ranging 168
Figure 7–13: Transition to more robust burst profile 170
Figure 7–14: Transition to less robust burst profile 170
Figure 7–15: DSA message flow—SS initiated 171
Figure 7–16: Provisioned authorization model 174
Figure 7–17: Dynamic authorization model 175
Figure 7–18: Two terminals with same planned PHY mode 189
Figure 7–19: Clock comparison 202
Figure 7–20: ARQ block states at the transmitter 208
Figure 7–21: ARQ block reception 211
Figure 7–22: Transmitter-initiated reset 214
Figure 7–23: Receiver-initiated reset 215
Figure 8–1: Encrypted payload format in AES-CCM mode 222
Figure 8–2: Initial CBC block and nonce 223

Figure 8–3: Counter blocks 223
Figure 8–4: AK management in BS and SS 225
Figure 8–5: TEK management in BS and SS 227
Figure 9–1: Three BSs required for full coverage 231
Figure 9–2: Two BSs required with mesh SS 231
Figure 9–3: Neighborhood definitions 233
Figure 9–4: Mesh CIDs 234
Figure 9–5: Mesh network entry: Opening a sponsor channel 238
Figure 9–6: Distributed scheduling 241
Figure 9–7: Centralized scheduling 242
Figure 9–8: Addition of PtP 245
Figure 9–9: Physical or directed mesh 246
Figure 10–1: TDD DL subframe structure 249
Figure 10–2: FDD DL subframe structure 251
Figure 10–3: DL MAP usage with shortened FEC blocks—TDM case . . . 253
Figure 10–4: Format of DL transmission CS PDU 254
Figure 10–5: Conceptual block diagram of DL PHY 255
Figure 10–6: UL subframe structure 257
Figure 10–7: Conceptual block diagram of UL PHY 258
Figure 11–1: OFDM symbol structure 263
Figure 11–2: OFDM frequency description 265
Figure 11–3: Subchannel subcarrier allocations 268
Figure 11–4: DL and network entry preamble structure 271
Figure 11–5: Example OFDM structure with TDD 273
Figure 11–6: DL burst and MPDU mapping 274
Figure 11–7: Mesh frame structure 276
Figure 11–8: RS-CC encoding process 279
Figure 11–9: BTC encoding block 280
Figure 11–10: CTC encoder 280
Figure 11–11: Initial ranging formats 283
Figure 12–1: Schematic indication of logical mapping elements 291
Figure 12–2: FUSC mapping example 295
Figure 12–3: PUSC DL mapping example 297
Figure 12–4: PUSC UL mapping example 299
Figure 12–5: Frame structure (TDD) 301
Figure 12–6: General shift index matrix structure 304
Figure 12–7: Single-slot initial ranging transmission for OFDMA 307
Figure 13–1: Example beam-forming network 312

Figure 13–2: TDD framing for AAS Systems 315
Figure 13–3: FDD framing for AAS systems 315
Figure 13–4: Open-loop transmit diversity schemes 321
Figure 13–5: Four-antenna STC and SM 324
Figure 14–1: WirelessMAN-OFDM throughput per 5 MHz channel 328
Figure 14–2: WirelessMAN-OFDM outage relative to 1.5 Mbit/s 329
Figure 14–3: Average throughput delay vs offered load 330
Figure 14–4: DL, IP traffic mean transfer delay 332
Figure 14–5: UL, IP traffic mean transfer delay 332
Figure 14–6: UL, Average transmit opportunities per frame 333
Figure 14–7: UL load and utilization of transmit opportunities 334
Figure 14–8: FCH + MAP overhead 334
Figure 14–9: UL, full-duplex SSs, VoIP mean transfer delay 335
Figure 14–10: UL, half-duplex SSs, VoIP mean transfer delay 336
Figure 14–11: DL, half-duplex SSs, HTTP mean transfer delay 337
Figure 14–12: UL, half-duplex SSs, HTTP mean transfer delay 337
Figure 14–13: UL, half-duplex SSs, HTTP mean transfer delay w/ARQ . . 338
Figure 14–14: OFDM CBLERs—ITU VehA 3 km/h 339
Figure 14–15: OFDM CBLERs—ITU VehA 30 km/h 340
Figure 14–16: Channel utilization as function of offered load 341
Figure 14–17: Modulation/coding utilization as function of offered load . . 341
Figure 14–18: Call blocking (C/I < 1 dB) as function of offered load 342
Figure 14–19: DL sector capacity per cell radius—ITU VehA (60 km/h) . 344
Figure 14–20: UL sector capacity per cell radius—ITU VehA (60 km/h) . 344
Figure 14–21: DL sector capacity per cell radius (ITU outdoor-indoor) . . 345
Figure 14–22: UL sector capacity per cell radius (ITU outdoor-indoor) . . 345
Figure 14–23: Delay for VoIP using rtPS, UGS, and ertPS 347
Figure 14–24: Average cell throughput for various schemes 347
Figure 14–25: CBLER, UL PUSC, ITU VehA 3 km/h, CC, ideal 349
Figure 14–26: CBLER, UL PUSC, ITU VehA 3 km/h, CC, estimated 350
Figure 14–27: CBLER, UL PUSC, ITU VehA 120 Km/h, CC, ideal 351
Figure 14–28: CBLER, UL PUSC, ITU VehA 120 km/h, CC, estimated . . 351
Figure 14–29: CBLER, UL PUSC, ITU VehA 3 km/h, CTC, ideal 352
Figure 14–30: CBLER, DL PUSC, ITU VehA 3 km/h, CC, ideal 353
Figure 14–31: CBLER, DL PUSC, ITU VehA 3 km/h, CC, ideal 354

List of Tables

Table 2–1: IEEE 802.16 members by geography, as of January 2006 21
Table 4–1: Key IEEE 802.16 features and their status 95
Table 6–1: CID allocation and well-known CIDs 113
Table 6–2: BW request header fields 115
Table 6–3: Generic MAC header fields 117
Table 6–4: Generic HT encodings 118
Table 6–5: MAC header demultiplexing 119
Table 6–6: FSH fields 121
Table 6–7: Fast-feedback allocation feedback type encodings 124
Table 6–8: Subheader and special payload ordering 127
Table 6–9: DSA-RSP message format 137
Table 6–10: Acknowledgment types 141
Table 6–11: RCID field interpretation 147
Table 7–1: Mapping of ATM CoS concepts to WirelessMAN 183
Table 7–2: ARQ parameters and timers 204
Table 12–1: Subchannel allocations 289
Table 14–1: Cell-ranges DL PUSC 354
Table 14–2: Relative performance benefit of MIMO and virtual MIMO . . 355
Table 14–3: Performance using MIMO 356
Table 15–1: WirelessMAN-SC system profiles 359
Table 16–1: Comparison of IEEE Std 802.11 and IEEE Std 802.16 372
Table 16–2: IEEE 802.20 performance targets 373
Table A–1: MAC headers, subheaders and special payloads 383
Table A–2: Management messages 384

Acronyms and abbreviations

The following acronyms and abbreviations are used in this book:

3DES	triple data encryption standard
AAS	adaptive antenna systems
ABR	available bit rate
ACID	ARQ channel identifier
ACK	acknowledgment
AES	advanced encryption standard
AISN	ARQ identifier sequence number
AK	authorization key
AMC	adaptive modulation and coding
AoA	angle of arrival
AoD	angle of departure
AP	access point
ARQ	automatic repeat request
ASIC	application-specific integrated circuit
ATDD	adaptive time division duplexing
ATM	asynchronous transfer mode
ATS	abstract test suite
BCC	block convolutional code
BE	best effort
BER	bit error rate
BPSK	binary phase shift keying
BRAN	Broadband Radio Access Networks (ETSI Technical Committee)
BRS	broadband radio service
BS	base station
BSN	block sequence number
BTC	block turbo code

BW	bandwidth
BWA	broadband wireless access
CAC	call admission control
CAZAC	constant amplitude zero autocorrelation
CBC	cipher block chaining
CBLER	coded block error rate
CBR	constant bit rate
CC	convolutional coding
CCM	counter with CBC-MAC
CDMA	code division multiple access
C/I	carrier to interference ratio
CID	connection identifier
CINR	carrier to interference-plus-noise ratio
CIR	channel impulse response
CLR	cell loss ratio
CP	cyclic prefix
CoS	class of service
CPE	customer premises equipment
CPS	common part sublayer
CRC	cyclic redundancy check
CS	convergence sublayer
CSCF	centralized scheduling configuration
CSCH	centralized scheduling
CSMA/CA	carrier sense multiple access/collision avoidance
CTC	convolutional turbo code
CTS	clear to send
DAMA	demand-assigned multiple access
DBPC	downlink burst profile change
DCD	downlink channel descriptor
DCF	distributed coordination function
DES	data encryption standard
DFS	dynamic frequency selection
DHCP	Dynamic Host Configuration Protocol

DIUC	downlink interval usage code
DL	downlink
DLFP	downlink frame prefix
DMZ	diversity MAP zone
DOCSIS®	data over cable service interface specification
DSA	dynamic service addition
DSC	dynamic service change
DSCH	distributed scheduling
DSCP	differentiated services code point
DSD	dynamic service deletion
DSL	digital subscriber line
DVB	digital video broadcast
EAP	Extensible Authentication Protocol
EDCA	enhanced distributed channel access
EGC	equal gain combining
EIRP	effective isotropic radiated power
EKS	encryption key sequence
FC	fragment control [field]
FCH	frame control header
FDD	frequency division duplexing
FDE	frequency-domain equalizer
FDM	frequency division multiplexing
FDMA	frequency division multiple access
FEC	forward error correction
FFSH	fast-feedback allocation subheader
FFT	fast Fourier transform
FHDC	frequency-hopped diversity coding
FHSS	frequency-hopping spread spectrum
FIFO	first-in, first-out
FSDD	frequency shift division duplexing
FSH	fragmentation subheader
FSN	fragment sequence number
FTP	File Transfer Protocol

FUSC	full usage of subchannels
GBN	go-back-n
GF	Galois field
GFR	guaranteed frame rate
GMSH	grant management subheader
GP	guard period
GPS	global positioning system
GSM	Global System for Mobile Communications
HARQ	hybrid automatic repeat request
HCCA	HCF coordinated channel access
HCF	hybrid coordination function
HCS	header check sequence
H-FDD	half-duplex frequency division duplexing
HSDPA	high-speed downlink packet access
HT	header type
HTTP	Hypertext Transfer Protocol
I	in-phase
ID	identifier
IE	information element
IFFT	inverse fast Fourier transform
IP	Internet Protocol
ISI	intersymbol interference
IUC	interval usage code
IV	initialization vector
KEK	key encryption key
LAN	local area network
LDPCC	low-density parity check coding
LLC	logical link control
LMDS	local multipoint distribution service
LMSC	LAN/MAN Standards Committee
LOS	line-of-sight
LSB	least significant bit
MAC	medium access control layer

MAN	metropolitan area network
MBS	maximum burst size
MIB	management information base
MIMO	multiple-input, multiple-output
MMDS	multichannel multipoint distribution service
MPDU	MAC protocol data unit
MPLS	multiprotocol label switching
MRC	maximum ratio combining
MSB	most significant bit
MSDU	MAC service data unit
MSH	mesh subheader
MS	mobile station
NACK	negative acknowledgment
NLOS	non-line-of-sight
nrtPS	nonreal-time polling service
nrtVBR	nonreal-time variable bit rate
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
O-FUSC	optional full usage of subchannels
O-PUSC	optional partial usage of subchannels
OSI	Open System Interconnection
PAN	personal area network
PAPR	peak to average power ratio
PAR	project authorization request
PCF	point coordination function
PCR	peak cell rate
PCS	packet convergence sublayer
PDU	protocol data unit
PER	packet error rate
PHS	payload header suppression
PHSF	payload header suppression field
PHSI	payload header suppression index
PHSM	payload header suppression mask

PHSS	payload header suppression size
PHSV	payload header suppression valid
PHY	physical layer
PICS	protocol implementation conformance statement
PKM	privacy key management
PM	poll-me
PMP	point-to-multipoint
PN	packet number
PoS	point of sale
PRBS	pseudo-random binary sequence
PS	physical slot
PSH	packing subheader
PtP	point-to-point
PUSC	partial usage of subchannels
Q	quadrature
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quadrature phase-shift keying
RCID	reduced connection identifier
RCT	radio conformance test
RF	radio frequency
RLC	radio link control
RS	Reed-Solomon
RS-CC	Reed-Solomon concatenated with convolutional coding
RSSI	received signal strength indicator
RSV or Rsv	reserved
RTG	receive-transmit transition gap
rtPS	real-time polling service
RTS	request to send
rtVBR	real-time variable bit rate
SA	security association
SAID	security association identifier
SAP	service access point

SC	single carrier
SCR	sustained cell rate
SDMA	spatial division multiple access
SDO	standards developing organization
SDU	service data unit
SFID	service flow identifier
SHA-1	Secure Hash Algorithm 1
SI	slip indicator
S/I	signal to interference ratio
SISO	single-input, single-output
SLA	service-level agreement
SM	spatial multiplexing
SNMP	Simple Network Management Protocol
SNR	signal to noise ratio
SOHO	small office/home office
SPID	subpacket identifier
SR	selective repeat
SS	subscriber station
SSRTG	subscriber station receive-transmit transition gap
SSTG	subscriber station transition gap
SSTTG	subscriber station transmit-receive transition gap
STBC	space-time block coding
STC	space-time coding
STTD	space-time transmit diversity
TCM	trellis-coded modulation
TCP	Transmission Control Protocol
TDD	time division duplexing
TDM	time division multiplexing
TDMA	time division multiple access
TEK	traffic encryption key
TFTP	Trivial File Transfer Protocol
TLV	Type-Length-Value
ToS	type of service

TSS&TP	test suite structure and test purposes
TTG	transmit-receive transition gap
UBR	unspecified bit rate
UCD	uplink channel descriptor
UDP	User Datagram Protocol
UGS	unsolicited grant service
UIUC	uplink interval usage code
UL	uplink
VC	virtual channel
VCI	virtual channel identifier
VLAN	virtual local area network
VoIP	voice over Internet Protocol
VP	virtual path
VPI	virtual path identifier
Wi-Fi®	wireless fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	wireless local area network
WPAN	wireless personal area network
WRAN	wireless regional area network

Chapter 1 Broadband wireless access (BWA)

Applicable market segments and requirements

The IEEE 802.16 Working Group on Broadband Wireless Access has been developing IEEE Std 802.16™ since 1999. Its original air interface standard was completed in 2001, but it has continued to evolve and will continue to evolve. The basic goal of the standard is to provide specific technologies and protocols to specify the air interface of BWA systems.

The central aim of IEEE 802.16 technology is to support broadband access; that is, to support access (i.e., connection to core networks) at broadband rates (which, according to a popular International Telecommunication Union (ITU) definition, means providing service at a rate of at least 1.544 Mbit/s, although we take the definition somewhat more loosely). Since IEEE 802.16 provides broadband access with a wireless connection, it could perhaps be said to provide "wireless broadband access." In the early days, the IEEE 802.16 Working Group considered this word order, finally settling on "broadband wireless access." It appears that this might emphasize that the work involved a broadband extension of the wireless access concept rather than a wireless implementation of broadband access concepts. In fact, because this standard does both, it exemplifies the ongoing trends toward convergence in telecommunications and data communications. The final choice of word order was rather arbitrary.

As the world's backbone telecommunications networks flourished in the twentieth century, vast data-carrying capacity is now available in fiber-optic lines. Unfortunately, this bandwidth is directly available only to those with direct access to the infrastructure, which is a very small fraction of the potential users. In the developed world, powerful fiber infrastructure may connect large commercial buildings but not the neighboring medium-sized commercial facilities. In many parts of the developing world, connection to

0022

the world's fiber infrastructure is still only a plan, or perhaps only a dream. When such a connection does take place, it will initially be at a single port of entry for an entire country. Distribution is critical.

Broadband access is about bridging the gap between the core infrastructure networks and the user's networks. This bridge can itself be built with wires or fibers, but the basic structure of the problem means that this requires a massive proliferation of cabling in order to serve users in a wide variety of places. This is a fundamental barrier to cabling. It is costly to install and maintain. Cabling hung from utility poles requires a series of permissions and rights-of-way. It is costly to deploy and subject to damage caused by everything from weather to traffic accidents. Buried cabling is more reliable (although still subject to cuts), but trenching is expensive and slow, or even prohibitive, due to the disruption to communities and traffic patterns. Doug Lockie, a veteran of the BWA business, has noted many times that "backhoes don't follow Moore's Law." His point is that no amount of exponential improvement in cost and performance of network equipment will enable cost-effective network deployment as long as cable installation is the fundamental limitation.

If broadband access is wireless, many of these construction problems can be bypassed. Compared to wired broadband access, wireless broadband access leans more heavily on technological innovation and less heavily on the physical plant. This is one reason it seems to hold more promise for the future. Of course, the wireless medium has its own limitations and costs; these are associated primarily with spectrum access, restrictions, and limitations. It remains to be seen how the costs and benefits of the wired and wireless cases will balance. It seems clear that cabling will be the leader at the core network side, where data are massively consolidated, and wireless will shine nearer the user, where data are more user-specific. The details of which technologies will be most competitive in various market segments will depend on many details, including the local infrastructure and regulatory environment.

Competition, or course, is driven by demand, and demand cannot always be satisfied by a single technical solution. If all homes and enterprises in an area have direct access to wired broadband, many of the consumers will probably be satisfied. On the other hand, some might still be willing to consider

alternative providers based on price, performance, features, or reliability; therefore, alternatives such as wireless broadband access may see an opportunity. But demand in the telecommunications world can also come from another direction: from the top down. Namely, many operators of core networks intend to provide service to retail customers. If these operators lack access networks, they are forced into leasing or partnership arrangements with the access provider. In an area with only one access provider (or only very few), the network operator may be forced into an unfavorable position. In such cases, demand for alternative access may be driven in top-down fashion by the operators.

In many cases, governments are seeking to foster more players and more competition in the broadband network marketplace. The market is sometimes slow to respond to demand, frequently because of natural barriers. For instance, the wired access market is difficult to penetrate, not only because it is slow and costly but, as noted earlier, because the development of local wired access facilities is burdened by regulation, especially because it is disruptive to environments (such as, for example, cities that cannot tolerate streets being trenched, or rural areas in which power lines affect wildlife migration). Barriers to wireless access networks are generally less prohibitive; therefore, wireless networks can be erected relatively inexpensively and also quite quickly. This, of course, depends heavily on spectrum availability and conditions.

One aspect of cabled infrastructure that is hard to avoid is its stationary nature. It is difficult to imagine cable being so densely deployed that it is readily available everywhere in a region the size of a city, or larger. Cable operators pick and choose their locations. Wireless signals, on the other hand, have the ability to cover a metropolitan-sized (or larger) area quite densely so that signals are broadly available, even to moving users. This allows for usage models that simply cannot be accommodated with cabled systems. The great technological and marketplace achievements of the mobile telephone business have helped to demonstrate the appeal of such a system and the flexibility it offers. Meanwhile, the success of portable computers, especially those networked with wireless local area networks (WLANs) (based on IEEE Std 802.11™) have demonstrated the appeal of untethered broadband access. This background points to a future in which broadband access moves toward

an increasingly portable, nomadic, and mobile usage model. The possibility of this evolution strengthens the hand of wireless broadband access relative to its wired alternatives.

As already noted, IEEE Std 802.16 is an evolutionary standard. The standard was designed originally for the support of stationary, enterprise-class deployments. However, the long-term goal was always to evolve the standard along with the developing technology to the point at which it could be economically feasible to move deeper into the access network and closer to the user. As of 2003, the standard was first enhanced with technology suitable for residential-class applications. Around the same time, work began to evolve the technology further, toward systems that could support mobile as well as stationary terminals. The IEEE 802.16e amendment, approved in December 2005, brings support for mobile as well as fixed terminals. We do not significantly discuss IEEE Std 802.16e™ in this book. However, in the standard's evolution, it continues to build on the foundations that were previously laid. Therefore, understanding the stationary ("fixed") technology is an essential basis for understanding the fixed/mobile advances.

As IEEE Std 802.16 has grown (see [B13][1] and [B38]), much of the evolution has occurred in the physical layer (PHY), which is the primary arbiter of the physical environment in which the technology can operate. All of this development, however, is based on the essence of the standard: its medium access control layer (MAC) specification. This MAC, which supports all of the standard's PHY options, was originally designed for the first application: carrier-quality, enterprise-based telecommunications services. Because of this history, the IEEE 802.16 standard can support the most demanding service requirements, even as it evolves toward more consumer-friendly applications. This is a key factor in the applicability of the standard. One of the primary MAC-related features of the standard's technology is its support for differentiated quality of service (QoS) among its users. An IEEE 802.16 base station (BS) can simultaneously support a variety of customer service requirements, for example, very demanding services such as real-time video

[1] Numbers in brackets refer the reader to additional resources listed in the bibliography of this book.

conferencing along with T1/E1 service, voice over Internet Protocol (VoIP), and best-effort Internet.

Spectrum, as the wireless medium, is a very precious infrastructure commodity. Any standard intended to support wireless communications, in any application, must be critically conservative in its use of this resource. IEEE 802.16 takes pains to conserve spectrum and emphasize spectral efficiency, at the expense of protocol complexity. This approach generally enhances its applicability, although perhaps some applications might be better suited to lower complexity protocols with lower efficiency.

IEEE Std 802.16 is called *the WirelessMAN standard for wireless metropolitan area networks*. "Metropolitan" in this sense indicates not the target geography but instead the target scale. The standard supports networks that are about the size of a city. It is by no means limited to urban applications. Some of the most likely applications are in rural areas in which high-quality broadband access is not readily available. The term *metropolitan area network* (MAN) predates the term *WirelessMAN* by many years. IEEE 802®, under which IEEE 802.16 is chartered, is formally known as the "LAN/MAN Standards Committee." It was called the "LAN Standards Committee" when formed in 1980, but the name grew with the group's portfolio.

IEEE 802.16 is an air interface standard, not a manual for service deployment. It is intended to support transport of any higher layer network requirement and in any application in which someone sees fit to apply it. The goal is flexibility. The developers of the standard may have had some applications in mind, but a good, open standard can be put to a variety of users by people who did not develop it. If a developing area is seeking to install a unified wireless network for voice, data, and video services, thereby obviating the need for a multiplicity of more narrowly focused systems as would be seen in a developed city, then IEEE 802.16 may be a good choice. Likewise, if a city is seeking to enhance the availability of broadband services to its population by constructing a public wireless data network, then IEEE 802.16 may also be an excellent candidate, serving either as a backhaul network supporting the ready deployment of WLAN access points (APs) or as an access network bringing a connection directly to a user. Some operators may seek to provide service to a fixed antenna, either on the outside of a building or indoors. Other operators

may seek to communicate directly with portable computers with on-board IEEE 802.16 radios.

In some cases, IEEE Std 802.16 might be a pure consumer system, with no commercial operator. Such use could be, for example, purely inside a home or enterprise, where the standard's QoS support gives it a leg up on alternative technologies. Or perhaps hobbyist or quasi-professional users will set up license-exempt systems for neighborhood communications. The applications are limitless, as they were intended to be.

Given this flexibility, it is difficult to completely address the specific market applications to which IEEE Std 802.16 may be applied. An amalgam of typical deployment scenarios is sketched in Figure 1-1. Many services are supported, consumer as well as commercial, mobile as well as fixed, and backhaul-only as well as directly to the end user. Such a mixed set of services signifies a mature and successful MAN. Perhaps we shall someday see such a network based on IEEE Std 802.16.

Below, we continue this introductory chapter, providing examples of how IEEE Std 802.16 might be applied to some idealized market situations.

COMMERCIAL FIXED BROADBAND WIRELESS: FIBER EXTENSION

In large metropolitan areas throughout much of the developed world, commercial office towers are connected to core networks by high-capacity fiber optic links, with broadband network services provided to the tenants. In the meantime, other smaller businesses, even those located across the street or on the next block, go without proper network facilities for a lack of fiber connectivity. IEEE 802.16 networks can fairly easily extend the reach of the fiber links. Suitable spectrum is available for this purpose in many areas. For example, the United States licensed 1150 MHz of spectrum in the 28–31 GHz range in 1998 for a service called *local multipoint distribution service* (LMDS); since then, additional higher frequency spectrum has become available on a license-exempt basis or with a low-cost license. Supply for spectrum at such frequencies outweighs demand primarily because of the less favorable propagation properties of electromagnetic radiation in this band. In particular, attenuation in air is rather high so that propagation distances are

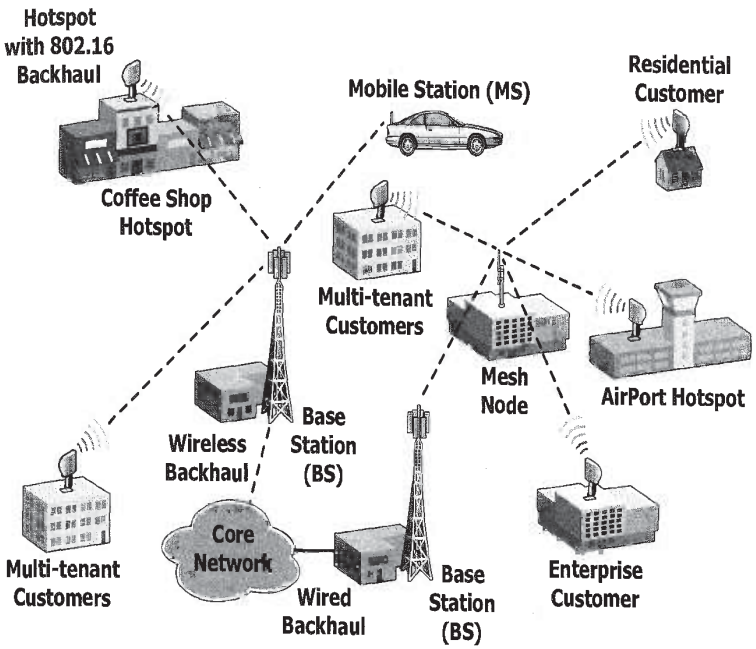


Figure 1–1: Typical IEEE 802.16 deployment scenarios

typically limited to around 3 km to 5 km. Also, attenuation increases dramatically in the presence of rainfall so that power control over a significant range is required. Diffraction is limited so that only line-of-sight (LOS) propagation is practical for communications. Furthermore, electronic components at these frequencies are relatively expensive. On the other hand, spectrum is plentiful enough for massive bandwidth, and short propagation distances and LOS propagation (supported by the compact, high-gain directional antennas available at these short wavelengths) allow for high frequency reuse.

Tall towers, which often have fiber connectivity, are good platforms for LOS BS antennas because the view from the top typically includes hundreds of

prospective customers. Commercial buildings allow for rooftop antennas, the cost of which (both the hardware and the professional installation) is spread over a number of commercial customers. In LOS links, multipath propagation is usually minimal; therefore, relatively simple modulation schemes may be effective. The extension of fiber requires the ability to support high data rates (> 100 Mbit/s), with high reliability and service-level guarantees. The ability to transport voice, video, and data traffic with appropriate quality is also an important requirement.

The first version of IEEE Std 802.16, approved in 2001, included a PHY specification, called *single-carrier WirelessMAN* (WirelessMAN-SC), suitable for use in the high-frequency bands associated with fiber extension. At the time, prospects for the deployment of the standard seemed rosy. However, the bursting of the "telecom bubble" early in the twenty-first century put an end to many promising plans. Currently, conditions are becoming more favorable for reconsideration of these applications. If so, then the WirelessMAN-SC air interface will be a good candidate to support such deployments. It supports the requirements, and it allows for stepped investment commensurate with market capacity requirements.

RESIDENTIAL FIXED BROADBAND WIRELESS: DIGITAL SUBSCRIBER LINE (DSL) AND CABLE MODEM ALTERNATIVE

In the developed world, broadband access has become a virtual necessity in the residential market. Homes are frequently the site of full-time or off-duty business, and broadband Internet access is in demand for education, information, voice, video, shopping, and entertainment. Cable modem and DSL networks have become the primary technologies addressing these home markets in many countries. However, such networks do not extend everywhere. Not all areas are served by cable television networks, and not all have been upgraded for cable modem service. While telephone networks are widely deployed in the developed world, not all lines have been upgraded for DSL service, and many homes are too far from the central office for DSL transmission.

Where DSL and/or cable modem service is available, BWA may nevertheless play a role. Customers may be seeking alternatives based on price,

performance, and/or reliability. Also, wireless service may have advantages, such as, in some cases, the ability to move a portable terminal throughout a service area. As previously noted, network operators who are dissatisfied with their inability to directly access their retail customers except through a local access provider may also be motivated to deploy a wireless access alternative.

In the developing world, and in many rural areas of the developed world, broadband access is simply unavailable. Wireless access may be particularly attractive in sparsely populated rural areas in which the cost of cabling is prohibitive.

Different design considerations may apply in rural and urban areas. For instance, in rural areas, economic considerations may require that a BS cell size be quite large, up to tens of kilometers. In these cases, LOS propagation with rooftop-mounted directional antennas may be appropriate. On the other hand, urban and suburban applications typically assume non-line-of-sight (NLOS) operation due to the proliferation of obstacles, the high incidence of multipath propagation, and the carrier frequency used (typically under 11 GHz). These networks may use licensed or license-exempt bands. Attractive licensed frequencies are available in the United States in the broadband radio service (BRS) bands [formerly known as multichannel multipoint distribution service (MMDS)] at around 2.5 GHz to 2.69 GHz. Outside the United States, other licensed bands, such as 3.5 GHz, are more common.

License-exempt operation has some additional challenges in MANs. Popular WLANs, which similarly operate in license-exempt 2.4 GHz and 5 GHz bands, are typically deployed in buildings and other areas under the control of a single entity deploying the network. This controlled environment is similar to a licensed deployment, since interference is limited to neighbors normally isolated to some degree by space and walls. However, license-exempt outdoor BWA deployments bring a different set of challenges, as a single entity cannot enforce its rules on other legal transmitters that may interfere. Any BWA solution has to take this into account.

In all of these cases, operators favor solutions that allow self-installation of customer equipment, since professional installation significantly affects the cost. Ideally, the customer equipment is in the form of a portable device that

0026

26

can be placed indoors. Such installation not only reduces the deployment cost but also offers the advantage of portability.

With residential customers becoming more technologically demanding, a competitive residential service offering is becoming suitable to small businesses as well. Because IEEE Std 802.16 is designed to support mixed services, it has no problem supporting such a customer mix. We can easily envision a multitude of applications for such a system. For instance, T1/E1 customers could be provisioned, and commercial WLAN hot spots, which typically require backhaul by 10BaseT (often over DSL), could be made entirely wireless using IEEE 802.16 backhaul. This would allow greatly increased flexibility in the placement of the IEEE 802.11 APs. Backhaul of other devices, such as cellular BSs, is another likely service.

IEEE Std 802.16 added support for a PHY at frequencies below 11 GHz beginning in 2003 with IEEE Std 802.16a™. This work, further refined in IEEE Std 802.16-2004 [B20] (the most recent revision of the standard), includes the specification of three PHY alternatives, known as orthogonal frequency division multiplexing (WirelessMAN-OFDM), orthogonal frequency division multiple access (WirelessMAN-OFMDA), and single carrier below 11 GHz (WirelessMAN-SCa). The first two, in particular, are well suited to NLOS operation in a multipath environment.

While IEEE Std 802.16-2004 specifies a standard for fixed access, the actual applications may allow a significant level of flexibility. For example, the user device may be *nomadic*, meaning that it can move as long as it does not operate while doing so. Such operation, as would be typical of a portable terminal such as a laptop computer equipped with wireless fidelity (Wi-Fi®), is well within the target market of deployments based on IEEE Std 802.16-2004. Even a certain amount of movement during operation is tolerable. Intercell handover, however, is beyond the scope of IEEE Std 802.16-2004. For this, users should turn to the IEEE 802.16e specification that was approved by IEEE as an amendment in December 2005.

At the time of publication, a number of companies have announced products claiming compliance to IEEE Std 802.16-2004.

QUALITY OF SERVICE (QoS)

In order to be competitive in a modern networking environment in which a single BS must support a multiplicity of widely varying transport demands over an inherently fluctuating wireless medium, a BWA network must include rigorous support for differentiated QoS as a fundamental design feature.

This requires support for flexible MAC and PHY framing to optimally use the available airtime adaptively. A centrally controlled MAC is also a necessity for BWA systems because distributed access methods, such as carrier sense multiple access (CSMA), cannot work efficiently in environments where some user devices are unable to hear other ones, due to distance, directional antennas, intervening terrain, etc.

The higher probability of errors in wireless environments introduces another challenge to the wireless QoS problem, especially in NLOS environments. The impact of errors on QoS can be minimized with sophisticated adaptive modulation and coding, error correction, error detection, and retransmission algorithms at both the MAC and PHY.

It is unnecessary, and in fact inappropriate, for a standard to enforce a specific QoS scheduling algorithm. However, a practical standard must define scheduling behaviors that allow the system to enforce uniform QoS. IEEE Std 802.16 does so by use of techniques analogous to those of the asynchronous transfer mode (ATM), where different scheduling behaviors are defined without specifying any particular scheduling algorithm.

The ability to provide QoS in both directions is also a critical requirement for efficiently managing the available spectrum and to support bidirectional applications such as voice and video conferencing.

THROUGHPUT REQUIREMENTS

Apart from basic QoS and reliability requirements, BWA systems must provide support a sufficiently high data transmission rate in order to be commercially successful. For backhaul systems, these throughput requirements are generally higher; depending on the load, systems should be able to carry a significant fraction of a fiber optic network's traffic. Moreover,

given the nature of the aggregated traffic over backhaul networks, the throughput requirements may be symmetrical [i.e., identical in both uplink (UP) and downlink (DL) directions], depending on the configuration.

On the other hand, the last-mile access or edge networks have to provide throughput comparable to that of competing technologies, such as DSL or cable. If BWA is used to support business needs, then the networks should be capable of supporting multiple business-class links such as T1, T3, DS3, or OC-3. Most importantly, the ability of the BWA network to scale efficiently with reasonable per-subscriber throughput is a critical requirement.

The increasing popularity of various multimedia, on-demand, and interactive applications is likely to increase the demand for network throughput and scalability in all application scenarios. IEEE Std 802.16 foresees the need for this evolution and is designed to support it.

Chapter 2 IEEE 802.16 standards

The working group and documents

BACKGROUND

IEEE Std 802.16, along with related standards and amendments, is developed and maintained by the IEEE 802.16 Working Group on Broadband Wireless Access. In this chapter, we begin with an overview of the umbrella organizations under which the IEEE 802.16 Working Group performs its activities, and we explain the basic process of developing and maintaining IEEE standards. We follow with a history of the IEEE 802.16 Working Group and review the historical development of its projects.

IEEE STANDARDS ASSOCIATION (IEEE-SA)

The Institute of Electrical and Electronics Engineers (IEEE) <http://ieee.org> is a technical professional society with over 350,000 members worldwide. IEEE has many technical and regional activities, most of which take place in a largely independent fashion.

The development of standards in IEEE is assigned to the IEEE-SA <http://standards.ieee.org>. The business of IEEE-SA is directed by an elected board of governors. The development and maintenance of standards are overseen by the IEEE-SA Standards Board, which mandates the process, approves the initiation of new projects, and approves appropriately balloted drafts as IEEE standards. IEEE-SA operates in accordance with the principles of consensus, due process, and openness defined by the American National Standards Institute (ANSI) and the Code of Good Practice for the Preparation, Adoption and Application of Standards produced by the World Trade Organization (WTO) under its Agreement on Technical Barriers to Trade. IEEE-SA is recognized by important international organizations as an international developer of standards, and IEEE standards are, in many cases, recognized as international standards. One example of particular relevance to

IEEE Std 802.16 is IEEE's international Sector Member status in the ITU's Radio Communication Sector (ITU-R), the same status held by the International Organization for Standardization (ISO).

The IEEE-SA leadership sets policy that directly influences not only the development of IEEE-SA standards but also their use. One critical topic is patents. The IEEE-SA patent policy is similar to that of most of the world's other formal standards developing organizations (SDOs). The key statement of IEEE-SA policy on this issue is "IEEE standards may include the known use of essential patents and patent applications provided the IEEE receives assurance from the patent holder or applicant with respect to patents whose infringement is, or in the case of patent applications, potential future infringement the applicant asserts will be, unavoidable in a compliant implementation of either mandatory or optional portions of the standard." The policy goes on to explain that, if the patent will be enforced, this "assurance" shall be "a statement that a license will be made available without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination." The IEEE generally avoids offering interpretations of these somewhat ambiguous statements, and any disagreements need to be settled outside of IEEE processes.

IEEE-SA standards are openly developed with consensus in mind. Participation in their development is entirely voluntary, as is their use. However, history has shown that standards developed in an open forum can produce high-quality, broadly accepted results that can focus companies and forge industries.

IEEE-SA charters over 200 "sponsor" groups to oversee the development of specific standards projects. These are typically organized within one or more of IEEE's technical societies.

While the sponsor groups have significant leeway in how they organize their work and prepare draft standards, the IEEE-SA is particularly assertive in the conduct of the ballot process under which drafts are reviewed as part of their consideration as IEEE standards. Before a prospective standard can be considered for approval by the IEEE-SA Standards Board, it must be balloted in a formal process that is defined by IEEE-SA. This process is, rather confusingly, known as "sponsor ballot" although it would more appropriately

be called "IEEE-SA ballot." In any case, balloting is conducted in an open process using a "ballot group" of volunteer individuals. IEEE-SA members are invited to participate in all ballots, regardless of whether they participated in the development of the draft. Balloting is an iterative process in which comments (a polite word for a specific complaint) are solicited and addressed, after which an improved draft is "recirculated" for further comment. As IEEE 802 Executive Committee Member Emeritus Geoff Thompson likes to say, balloting is about "improving" the draft, not about "approving" the draft. More comments lead to a better result. The IEEE 802.16 Working Group seeks comments, and the IEEE 802.16 task groups are experienced at resolving them (up to 500, or 1000, or even 2000 in a weeklong session, if necessary). Provided that a competent, active, and sincere ballot group is engaged in the process, the outcome is a sound and reliable technical document that reflects a broad consensus. While the participants are individual human beings, they often bring with them the technical ideas of their home environments, including national, regional, and corporate viewpoints.

IEEE 802® LAN/MAN Standards Committee (LMSC)

The development of local area network (LAN) and MAN standards with IEEE-SA is assigned to the LMSC, informally known as *IEEE Project 802* or simply *IEEE 802* <http://ieee802.org>. One of the largest, most prolific, and most influential of the IEEE-SA sponsors, IEEE 802 has operated since 1980 under the IEEE Computer Society. It develops and maintains standards addressing the MAC and PHY, each of which fits under a common logical link control (LLC) layer. Taken together, these make up the two lowest layers of the Open System Interconnection (OSI) seven-layer model for data networks (see [B29]).

IEEE 802 oversees a panoply of network standards, using an internal structure based on working groups developing draft standards. IEEE 802's great successes include IEEE 802.3 Ethernet, IEEE 802.11 WLANs, and IEEE 802.15.1 Bluetooth™ personal area networks (PANs). Other significant, although ultimately less successful, projects have included token ring and token bus.

IEEE 802 is a large but tightly managed organization that meets in plenary session each March, July, and November, with recent attendance in the range of 1800 people. The organization is governed by an executive committee composed of the chairs of the active working groups and technical advisory groups (currently 11 individuals) and 7 additional officers.

STANDARDS DEVELOPMENT IN IEEE 802

The IEEE 802 process is designed for quick development of standards with broad consensus. The demand for consensus helps to ensure that standards are technically superior and meet market needs.

The development process in IEEE 802 follows the chronological steps outlined below. The process is overseen by the IEEE 802 Executive Committee and defined by a written set of rules and procedures.

Study group stage

When sufficient interest has been identified in a topic, IEEE 802 may establish a study group to investigate the problem and consider the interest and potential scope of a possible standardization project. Should a study group wish to pursue standardization, it must draft a project authorization request (PAR); this is a form by which all new IEEE-SA projects become authorized. Before the Executive Committee considers approving a PAR for submission to the IEEE-SA Standards Board, it requires a statement addressing IEEE 802's "five criteria for standards development." This statement must demonstrate that the potential standard has broad market potential, compatibility with other IEEE 802 standards, distinct identity within IEEE 802, technical feasibility, and economic feasibility.

Working group development of draft

The Executive Committee assigns each new project to an existing or new working group and charters that group to develop the standard. Technical decisions are made in the working group by vote of at least 75% of its members. Membership in IEEE 802 working groups belongs only to individual people, usually engineering professionals, and is established and

maintained by participation in sessions, according to specific rules. Nonmembers participate actively as well, often with significant influence.

The initial draft development method varies among working groups, but the typical process is to delegate a task group to the problem and issue a public call for contributions requesting documentary input. Eventually, the task group develops a first draft, either by adopting a complete contribution or by assembling a collections of inputs. This process is sometimes contentious, as the competing interests of different companies and technology interests are often reflected in meetings. However, it offers a good opportunity for new participants to come to understand the process and become comfortable in discovering commonality among a group with diverse interests. The fact that professional individuals, not companies, are the recognized entities helps to set a tone of collegiality. New participants continuously enter the process. They sometimes appear aggressive at first, only to discover that this approach can be ineffective. The process demands excellent communication and preparation as well as technical skills. Those most effective at furthering their causes are those who clearly state their intent, present well-documented arguments, and look for opportunities to unite with others who have compatible goals. The system is an excellent training ground for bringing out effective communications skills, and many of the most effective participants learn their skills the hard way.

Once a working group has adopted a draft, the process changes subtly but significantly. At this point, if all goes well, the interest of the participants begins to align with the common goals of improving and completing the draft. The typical process is driven by distributing the draft and requesting comments in the form of specific requests to make changes. A strict IEEE 802 policy, atypical in IEEE-SA, is the requirement of a formal ballot process, modeled after the IEEE-SA ballot, before a draft standard may be forwarded to IEEE-SA for sponsor ballot. This dual ballot process is a key factor in the quality control for which IEEE 802 has become known.

In this "working group letter ballot," as in sponsor ballot (both of which are paperless), any vote against the document must be accompanied by specific comments on what changes are required in order to make it acceptable to the voter. This process forces constructive suggestions of change and helps drive

the process to quick improvement. Members voting to approve, and nonmembers as well, are also solicited for suggestions. An approval rate of 75% is required for draft acceptance. However, changes made in response to comments, and negative comments that have not been accepted by the editorial team, must be "recirculated" for approval by the voters. This allows for additional reviews and additional improvement. Eventually, however, most ballots reach a terminal period in which a large consensus favors closure. The ballot cannot close until those voting negative have had their say and failed to attract significant support for their argument. The approval margin is typically much higher than 75% at closure, but it need not be.

Following approval in working group letter ballot, drafts are forwarded for IEEE-SA sponsor ballot. This is similar to a rerun of the working group letter ballot except that the ballot group is not restricted to members of the working group. When this ballot is complete, the draft, ballot results, and supporting documentation are forwarded for review by the IEEE-SA Standards Board's Review Committee (RevCom). RevCom's recommendation proceeds to the full board for final action.

Once IEEE standards are approved, they are professionally edited and generally published and offered for sale within about two months, depending on size and complexity and the extent of editing required.

Unique to IEEE 802 within IEEE-SA is the "Get IEEE 802" program <http://standards.ieee.org/getieee802> in which published standards are available for download without charge beginning six months after publication. The cost of this program is subsidized by the individual participants, through the session registration fee, along with a few corporate sponsors.

IEEE 802.16 Working Group: Overview

IEEE 802's WirelessMAN work takes place within the IEEE 802.16 Working Group on Broadband Wireless Access <http://WirelessMAN.org>. The working group is a unit of IEEE 802, which serves as sponsor of IEEE 802.16 projects (although, unique to IEEE 802, the IEEE 802.16 projects also have a cosponsor: the IEEE Microwave Theory and Techniques Society).

IEEE 802.16 Working Group: History

The activities of the IEEE 802.16 Working Group were initiated by Roger Marks of the (U.S.) National Institute of Standards and Technology (NIST), who organized a meeting on BWA standardization, attended by 45 people, in August 1998 at the IEEE Radio and Wireless Conference in Colorado Springs, Colorado, USA. Following his visit to the IEEE 802 plenary session in July, Marks forwarded an invitation from IEEE 802 Chair Jim Carlo to convene a meeting on this topic at the IEEE 802 plenary session in November. At that session, IEEE 802 approved the formation of the Study Group on Broadband Wireless Access. That study group met twice and drafted a PAR, limited to 10 GHz to 66 GHz, that was endorsed by the IEEE 802 Executive Committee in March 1999. This action (after approval from the IEEE-SA Standards Board) created the IEEE 802.16 Working Group. Following an organization session (Session #0) in May, 106 people became charter members of the working group at its first official session in Montreal, Canada, in July 1999.

The working group has continued to meet at each IEEE 802 plenary session (in March, July, and November) and to hold a working group interim session each January, May, and September. (An additional interim session in August 1999 caused the even/odd numbering of sessions to reverse; since then, even-numbered sessions correspond to IEEE 802 plenaries.) As shown in Figure 2–1, which includes sessions through #39 of September 2005, attendance has grown and waned. Participation interest depends on the current activity. Interest was very high in 2004-2005 as the IEEE 802.16e amendment was being assembled, and attendance peaked at 367 in November 2004.

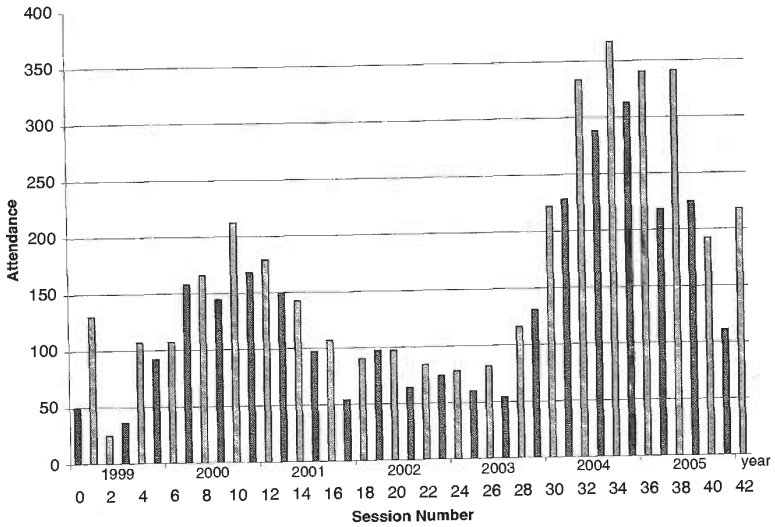


Figure 2–1: IEEE 802.16 Working Group attendance

Beginning in 2004, in recognition of the fact that IEEE 802 plenaries are generally in North America, nearly all sites selected by the working group for its interim sessions have been outside North America, most often in Asia. Like attendance, membership in the working group also fluctuates, lagging behind the attendance figures. As of November 2005, the working group had 310 individual members. According to the addresses they provided, they represent a broad geographical base, as shown in Table 2–1.

Table 2–1: IEEE 802.16 members by geography, as of January 2006

Address	Number of members
Canada	20
China	17
Finland	2
France	3
Germany	7
Ireland	1
Israel	18
Italy	3
Japan	7
Korea	60
Netherlands	4
Romania	1
Singapore	1
Sweden	3
Taiwan	3
UK	11
USA	149

TECHNICAL PROGRESS IN IEEE 802.16 WORKING GROUP

Since 1999, the IEEE 802.16 Working Group has constantly been active in developing standards projects, usually with multiple parallel activities. Although some of its projects have been very large, the group prefers to divide its efforts into specific problems that can be reasonably well-defined and completed within a predictable time. The IEEE process allows for the development of amendments that modify an existing standard. The published

amendment, which is designated by a lowercase letter after the primary standard number, is not an independent specification because it includes only the modifications, not the base material from the original standard. Upon approval of the amendment, the applicable standard is no longer the prior version, but the version defined by the application of the amendment. When appropriate, a revision of the standard may be undertaken; in this case, the base standard and its published amendments are editorially merged and reballoted, with the entire document open to comment.

Figure 2–2 shows a timeline of the past and current projects of the IEEE 802.16 Working Group. The designated start date is that of the approval of the project authorization (PAR); in some cases, the chart shows the date of a previous PAR that was later modified before the project was complete. The end date in each case is the actual or anticipated date of approval.

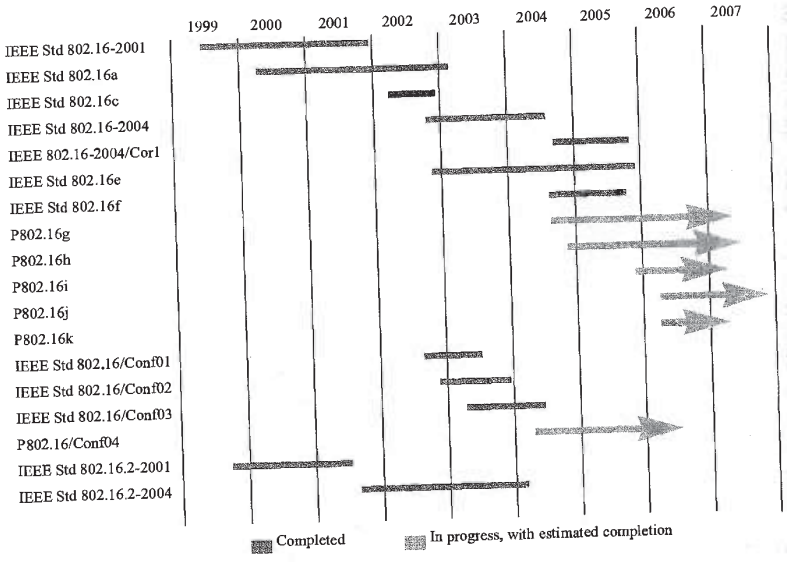


Figure 2–2: IEEE 802.16 project timeline

Air interface: IEEE Std 802.16

Work under the original IEEE 802.16 PAR, to develop an air interface for 10 GHZ to 66 GHz, began in July 1999. By November 1999, 35 PHY and MAC proposals were considered. By March 2000, two consolidated MAC/PHY proposals were still under consideration. In May 2000, agreement was reached to merge these proposals. A working group letter ballot followed soon thereafter. IEEE Std 802.16-2001, IEEE Standard for Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed Broadband Wireless Access Systems, was approved in December 2001. The MAC protocol is fundamentally based on a time division multiplexing/time division multiple access (TDM/TDMA) protocol supporting time division duplexing (TDD), frequency division duplexing (FDD), and half-duplex frequency division duplexing (H-FDD). The PHY, entitled *WirelessMAN-SC*, is a single-carrier system assuming LOS propagation to fixed antenna terminals.

Because the working group believed that IEEE Std 802.16-2001 allowed too many options for easy interoperability testing, it opened up a follow-up amendment project, IEEE P802.16c, to define a set of profiles that would each define a set of options with sufficient specificity to allow for interoperability. With Ken Stanwood's leadership as task group chair, this project both opened and closed in 2002.

In November 1999, while the 35 MAC and PHY proposals were being discussed, IEEE 802.16 created a study group, under the leadership of Brian Kiernan, to develop a PAR for frequencies below 10 GHz. That PAR was approved in March 2000 and began the working group's activities in NLOS PHY technology.

The working group targeted its higher frequency PHY at carrier frequencies above 10 GHz and its lower frequency work at less than 11 GHz. This overlap simply reflected the fact that some available bands (in particular, 10.5 GHz) seemed suitable for either. In fact, this minor overlap was dwarfed, early on, by more fundamental issues regarding the relationship between these two approaches. Initially, many of those interested in the low-frequency applications favored basing the work on a new MAC, believing that the original MAC was too enterprise-centric and not sufficiently suited to basic residential applications. However, upon further discussion, the working group

made the virtually unanimous decision to build all of its PHY specifications upon a single, sophisticated MAC foundation. Another contentious issue was whether to pursue separate PHY projects for licensed and licensed-exempt bands. After starting in this direction, the working group later decided to merge those efforts. As a result of these changing decisions, a number of PARs were revised and renamed in the early years. Eventually, the project became amendment project IEEE P802.16a, and the amendment was approved by IEEE-SA in early 2003.

IEEE Std 802.16a includes three separate PHY specifications:

- WirelessMAN-SCa: single carrier
- WirelessMAN-OFDM: multicarrier with 256 subcarriers
- WirelessMAN-OFDMA: multicarrier with 2048 subcarriers

The debate leading to the decision to include all three modes was long and contentious. The eventual decision was not to everyone's liking, but compromise is a critical element of standardization. In attempting to develop a single standard for worldwide use, it was necessary to recognize a number of different worldwide needs. Since IEEE Std 802.16a was adopted, WirelessMAN-OFDM has become very popular. WirelessMAN-OFDMA, which was more forward-looking at the time of adoption, is increasingly appearing to be the choice of the future, particularly as IEEE Std 802.16 is evolving toward mobile systems. In the meantime, WirelessMAN-SCa has not gained significant industry interest.

Once IEEE Std 802.16a was complete, the working group opened the amendment project IEEE P802.16d. This was intended to parallel the profiles project IEEE Std 802.16c, but oriented toward the lower frequencies. The PAR also allowed for the correction of errors that inevitably plague complex standards such as IEEE Std 802.16. Gordon Antonello agreed to chair the project, which was intended to be short and sweet, like IEEE Std 802.16c™. However, the project quickly grew in complexity, as a number of proposals looked more like enhancements than error corrections. Eventually, the working group decided that it needed to convert the project from an amendment into a revision; this would editorially merge IEEE 802.16-2001, IEEE 802.16a, and IEEE 802.16c, opening the entire result to comments

regarding corrections and improvements. The amendment PAR IEEE 802.16d was abandoned in favor of a revision project IEEE 802.16-REVd, and Antonello ended up with a much larger task than he had expected. The work was finally approved in June 2004 and weighed in at nearly 900 pages. Because it was a revision, IEEE Std 802.16-2004 [B20] made IEEE Std 802.16-2001, IEEE Std 802.16a, and IEEE Std 802.16c obsolete. Some people incorrectly refer to this document as the "16d" standard, but letters are used to identify amendments, not revisions.

Once again, bugs and errors turned up. In September 2004, a PAR was approved for a new project to address them. In IEEE parlance, a project allowing corrections but prohibiting new features is called a *corrigendum.* Under the leadership of Jon Labs, the corrigendum IEEE Std 802.16-2004/Cor1 was completed in September 2005 and approved in November 2005. This document put to rest, for the near term, the definition of the IEEE 802.16 air interface for fixed wireless access.

However, a working group in motion seems to remain in motion. As the amendment project IEEE P802.16a was wrapping up in late 2002, the working group opened up a new PAR, IEEE P802.16e, to expand the IEEE 802.16 fixed access system into a combined fixed/mobile system, allowing a single BS to support both fixed and mobile terminals in licensed bands below 6 GHz. The amendment project IEEE P802.16e, chaired by Brian Kiernan, attracted great interest and a great many participants to the working group. The influx of people tended to keep the project unstable; therefore, many decisions were revisited again and again. However, the work did come to a conclusion, and the final draft, at 684 pages, was approved in December 2005. IEEE Std 802.16e amends IEEE Std 802.16. All of the three lower frequency PHY modes are supported, but the WirelessMAN-OFDMA mode is made "scalable" with the addition of new subcarrier counts: 128, 512, and 1024. The details of IEEE Std 802.16e and its content is beyond the scope of this book.

In March 2006, Brian Kiernan was awarded the IEEE-SA Standards Medallion "for steadfast and exemplary leadership of the Task Groups developing the IEEE 802.16a and 802.16e WirelessMAN standards

specifying wireless metropolitan area networks for fixed and mobile broadband wireless access systems."

As the working group's attention turned to mobility, it decided that network management would be an increasingly critical issue. The Network Management (NetMan) Task Group, chaired by Phil Barber, was initiated to address such concerns. The NetMan group took on two projects: IEEE P802.16f and IEEE P802.16g. IEEE Std 802.16f™, a management information base (MIB) for fixed systems, was approved in September 2005. IEEE P802.16g is a complex activity on "management plane procedures and services." Approval is not expected until 2007. In the meantime, the working group began planning in late 2005 to follow up IEEE Std 802.16f with a new MIB project for the mobile case. This was launched as IEEE P802.16i in December 2005.

The work to amend IEEE Std 802.16-2004 also continues with IEEE P802.16h, which is attempting to address the long-neglected problem of coexistence in license-exempt bands. The License-Exempt Task Group leading the effort is chaired by Mariana Goldhamer.

In July 2005, following a number of presentations and expressions of interest, the working group created the Mobile Multihop Relay Study Group to investigate the initiation of a new project. The existing standard specifies both the BS and the subscriber station (SS). The study group was chartered to consider the additional specification of a relay station, which would offer a valuable new tool to system operators. The study group, chaired by Mitsuo Nohara, proposed an IEEE P802.16j PAR. Following a tutorial on the topic at the IEEE 802 plenary session in March 2006, the PAR was approved later that month.

Conformance: IEEE 802.16/Conformance0*X*

The IEEE 802.16 Working Group, aware that air interface standards alone cannot specify conformance or interoperability, believes in the importance of conformance test documents. The working group has completed and published three stand-alone conformance test standards, all applicable to WirelessMAN-SC systems and all developed by the Conformance Task Group, chaired by Ken Stanwood. IEEE Std 802.16/Conformance01-2003

[B21], IEEE Std 802.16/Conformance02-2003 [B22], and IEEE Std 802.16/Conformance03-2004 [B23] address the protocol implementation conformance statement (PICS) proforma, test suite structure and test purposes (TSS&TP), and radio conformance tests (RCTs), respectively. Work on a PICS for the PHYs operating below 11 GHz is taking place under the IEEE P802.16/Conformance04 project, chaired by Gordon Antonello.

Coexistence: IEEE Std 802.16.2™

Beginning in 1999, the working group took note of the difficulties that would be faced by operators of systems in licensed bands due to co-channel and adjacent channel interference. It opened a PAR to create a stand-alone *recommended practice*, IEEE parlance for a standard that uses the verb "should" instead of "shall" in its normative statements. IEEE Std 802.16.2-2001, addressing the frequencies important to the WirelessMAN-SC PHY, was approved in 2001. A revision, to include the lower frequencies in IEEE Std 802.16a, was approved in 2004 as IEEE Std 802.16.2-2004 [B26]. Phil Whitehead chaired the Coexistence Task Group for both projects.

Chapter 3 Basic concepts and definitions

Wireless protocol and communication concepts

This chapter introduces some of the basic concepts and definitions related to wireless design, antennas, regulatory issues, MAC, and the PHY. This will form a basis for the more detailed discussions in the rest of the book. Among the countless books on these issues, we refer the reader to a few on the topic of broadband wireless or wireless in general (see [B4], [B7], and [B51]) and on networking (see [B6], [B33], [B42], [B50], and [B54]). These and other references offer a more thorough study of the specific concepts.

In this chapter, we use the term *wireless device* to refer generally to a radio communication terminal, whether a central controller such as a BS, a client SS, or a node in a multihop network. Note that, according to this definition, "wireless" devices may in fact be cabled to a network, such as a backhaul network on the BS side or a user LAN on the SS side, but we call the device "wireless" because it carries out radio communications over a wireless link. The communication between wireless devices is accomplished using protocols and techniques that require an antenna for transmission and reception of radio waves, a PHY for modulation and demodulation of signals, and a MAC for coordinating and controlling access to the medium (the "airwaves"). Other networking, management, and privacy components may be present in a wireless device to support additional services. IEEE Std 802.16 provides support for point-to-multipoint (PMP) systems centered around a coordinating BS and a number of SSs. It also supports, as an option, a multihop mesh mode in which a wireless device receives information and forwards it to another wireless device.

FREQUENCY BANDS

One of the basic parameters that characterizes a wireless system is the frequency band it uses, generally expressed in terms of megahertz (MHz) or gigahertz (GHz).

Channels

A frequency band is often split into multiple frequency *channels* to support independent communication activities. A channel is usually defined with a specific center frequency and occupied channel bandwidth. For example, a 2437 MHz channel with a 22 MHz bandwidth occupies the spectrum from 2426 MHz to 2448 MHz.

Licensed and license-exempt spectrum

Typically, the frequency spectrum in which a specific wireless device operates can be classified into *licensed* or *license-exempt* (or *unlicensed*) spectrum. The allocation and regulation of licensed and license-exempt spectrum are typically controlled by regulatory agencies, such as the Federal Communication Commission (FCC) in the United States.

In licensed bands, individual licensees are typically granted exclusive rights by the regulatory agency to a portion of the spectrum in a specific geographical area, sometimes in exchange for a fee. For example, most mobile telephone operators hold exclusive rights to the spectrum in which they operate; in many cases, the license rights were obtained at auction. Conversely, operation in license-exempt spectrum does not require explicit permission, although generally transmission is permitted only using communication devices that have been certified to meet regulatory specifications. In general, operation in license-exempt situations offers no legal protection against harmful interference.

The distinction between licensed and license-exempt spectrum is, in many cases, overly simplistic. For instance, in some cases, license-exempt devices are permitted to operate in the same bands, and in the same general geographic areas, as licensed devices, provided that the licensed devices are protected from interference in some way. In this case, it is the device

operation, not the spectrum, that is properly characterized as licensed or license-exempt.

A licensing model commonly referred to as *light licensing* is coming into use by some regulatory agencies for some frequency bands. In the light licensing model, transmission requires a license, but licenses are nonexclusive. Typically, the license holders are required to register transmitter details with the regulatory agency. The regulator may control subsequent licenses to avoid interference with previous license grants in the same location, or it may leave interference issues to be negotiated by the licensees. With the development of smarter and cognitive wireless devices, many regulators are hoping to move toward less spectrum regulation and let technology address the interference issues.

In most cases, wireless devices, whether in licensed and license-exempt operation, require legal certification by the governing regulatory agency. The certification usually requires that devices meet specific requirements regarding fundamental physical features, such as transmitter power, in-band and out-of-band emissions, and channel selection. In other cases, regulations may specify the type of modulation and detailed protocols, although such an approach is beginning to look anachronistic. Rules may depend on the application; for instance, rules for indoor and outdoor operation may differ.

One relatively new requirement for license-exempt operation in certain cases is *dynamic frequency selection* (DFS). DFS is used by a wireless device to dynamically detect any active transmitters (typically, high-power radar) on the channel and take corrective actions. The actions may include stopping transmission on the occupied channel and switching to another channel. DFS may also be enhanced to support more sophisticated coexistence protocols, including those based on detection of lower powered nonradar devices.

In virtually any regulatory domain, many radio frequencies are unavailable for private terrestrial wireless communications, licensed or not. Many of these bands are allocated to specific public use, such as military or public safety, or to other private services. For a detailed list of allocated bands and specific requirements, please refer to the appropriate regulatory agency in the area of interest.

Spectrum and standardization

While the regulatory agencies do not generally take responsibility for developing standards for wireless communications, wireless standards developers must take regulatory requirements into account and provide solutions that meet those demands, as well as technical and business requirements. In the case of license-exempt operation, protocol designers, including those involved in standardization, must include the support of extra measures for coexistence and interference mitigation. In a license-exempt band, no single entity controls the deployment of devices. Operating devices may well transmit within range of each other. Without defined coexistence mechanisms and interference mitigation techniques, harmful interference may occur. For example, the 2.4 GHz industrial, scientific, and medical (ISM) band is heavily used by IEEE 802.11 WLANs, IEEE 802.15 wireless personal area networks (WPANs, including Bluetooth), cordless telephones, etc. IEEE 802.16 devices are also allowed, provided that they follow the regulatory requirements. The basic rules of operation may provide some protection from interference, and specific protocols can offer additional protection to devices jointly observing those protocols.

Coexistence

In licensed spectrum, the licensee manages and controls the transmitters. This priority control over emitters in the licensed spectrum and geographic area, sometimes combined with coordination with licensees in adjacent geographic areas, adequately ensures coexistence for the emitters of the licensee. Coexistence ensurance hence may take among others the forms of emitter positioning and manual configuration as well as automated configuration through coexistence protocols among emitters. IEEE Std 802.16 addressed the issue of licensed coexistence in IEEE Std 802.16.2-2004 [B26].

On the contrary, license-exempt spectrum usually allows the operation of many incompatible wireless protocols. A network including a set of wireless devices following a specific protocol may successfully share the spectrum. A different protocol might work equally well. However, the two networks, when operating in proximity, may appear to each other as unmanageable and harmful interference. Standards developers with products intended for

license-exempt operations are increasingly aware of the need to coexist peacefully with devices based on different protocols. For example, IEEE 802 maintains the IEEE 802.19 Coexistence Technical Advisory Group to "develop and maintain policies defining the responsibilities of 802 standards developers to address issues of coexistence with existing standards and other standards under development." However, practical techniques for coexistence across protocol lines are still in their infancy.

TYPES OF WIRELESS NETWORKS

Fixed and mobile networks

Wireless devices may be classified based on whether the communicating devices are physically fixed at one location or allowed to move during a communication session. For example, the ITU defines fixed service as a "radio communication service between specified fixed points." In *fixed* wireless networks, wireless devices at both ends are stationary during operation. In *mobile* wireless networks, at least one of the devices can move during operation. An example of the mobile wireless network is the cellular telephone network in which the cellular BSs are fixed and the phones or handsets are mobile. Other types of mobile networks offer a range of mobility and support a variety of device velocities.

The regulatory requirements of fixed and mobile wireless networks differ for a variety of reasons. For example, since fixed networks are stationary and not carried by humans during operation, they may be allowed to use higher transmit power in certain regulatory domains. Similarly, mobile devices, especially handsets carried by humans, may have a different set of regulations with respect the transmit power and allowed emission levels. Both fixed and mobile networks exist in both licensed and license-exempt spectrum.

Apart from the regulatory requirements, fixed and mobile devices differ in technical design. Fixed networks may (in some cases) use high-power transmitters and involve high-gain, directional rooftop antennas with a clear LOS. This situation provides for a simple propagation channel. Although some fixed-system designs are based on more challenging links, to account for indoor antennas with minimal installation problems, the channel is

relatively static. Mobile wireless networks, on the other hand, may involve a handheld subscriber unit with a small, omnidirectional antenna; therefore, the channel may be quite poor. Furthermore, since the device may be in rapid motion, the system must accommodate a rapidly varying channel. On the PHY level, the system must deal with changes in the radio frequency (RF) signal that are the result of the mobile device's movement, such as large-scale fading and Doppler. These phenomena will be discussed later in this chapter.

Effective mobile communication requires optimization at both MAC and PHY levels. An important additional requirement, with both PHY and MAC ramifications, is fast handover. For example, a mobile device communicating with a fixed BS must establish communication with a second BS, without disconnecting from the first, as it goes out of the range of the first device. Handover usually involves support from components outside the MAC and PHY. However, the MAC must include basic support for handover, working in conjunction with other functions.

Also, fixed devices are often attached to the electrical power grid so that power efficiency is a relatively minor issue. In mobile devices, power use is a foremost factor.

Nomadic and portable networks

Nomadic service refers to a mode of operation in which at least one of the wireless devices in the system is transportable but, during operation, the devices are stationary. For example, most laptops with IEEE 802.11 WLAN connections are operated in nomadic mode. *Portability*, on the other hand, is a physical property of the device itself. Portable devices may be mobile or nomadic from the communication perspective.

WIRELESS NETWORK TOPOLOGIES

The topology of a wireless network depends on how the wireless devices are connected and the role played by each wireless device. When a wireless network consists of only two directly connected devices, the topology is referred to as a *point-to-point* (PtP) network. In a PtP network, the access methods and protocols can be simplified, as there is no need to support

multiple access. PtP networks are typically used to connect backhaul networks and are often fixed networks. The wireless devices at both ends of a PtP network usually have similar functionality and capabilities.

On the other hand, a *point-to-multipoint* (PMP) network consists of a single wireless device (a BS) communicating directly with more than one wireless client device (the SSs), thereby representing a star topology. The BS acting as a controller at the center of the network is responsible for performing additional functions to coordinate the actions of other devices. A PMP network requires a multiple access method. The communication from the controller device to other wireless devices is referred to as *downlink* (DL) or *downstream*, and the communication in the reverse direction is referred to as *uplink* (UL) or *upstream*. The DL and UL directions are also referred to as *forward link* and *reverse link*, respectively.

Both PtP and PMP wireless networks are a class of *single-hop* networks, in which the wireless protocol defines the communication across one wireless link. Although the wireless device on both ends may be connected to other wired or wireless networks to support distribution, aggregation, or forwarding services, these networks are considered independent. However, a third type of wireless topology, known as a *mesh*, is in the class of *multihop* networks. A device or node in a mesh network is usually connected to two or more other devices, and the communication needs to be coordinated for more than one hop for a mesh network to operate properly. Apart from basic access and connection management, the mesh networks also need to define additional functions such as routing, fault-tolerance, and interference mitigation. All the devices on a mesh network are expected to have similar capabilities and functions, although some devices may have additional functionality to perform special functions in a mesh network, such as a connection to another network or as an aggregation point.

While these three types of network topology have elements in common, there are some significant differences. A network and protocol designed for a PMP topology may be simplified and optimized to a PtP network. This is often the case in practice, as both networks are single-hop networks. However, the mesh networks require additional functionality that usually requires a different protocol or a set of extensions to support. Since the frame structures

of the PtP/PMP and mesh operation can be significantly different, these modes not always interoperable.

The primary benefits of the mesh networks are as follows:

- Extended range with an extra hop
- Increased NLOS coverage when one or more mesh nodes are added to go around obstacles
- Alternate paths in case of failures or performance degradation

Mesh also comes with some disadvantages, including the following:

- Increased delays introduced by multiple hops
- Increased complexity of protocols (i.e., MAC, routing, management, security)
- Increased complexity of planning of initial coverage (network seeding)

If not properly managed, the delays introduced by multihop wireless networks may be unacceptable for certain types of applications.

RF PROPAGATION

LOS and NLOS

Wireless systems are often classified on the basis of whether restricted to LOS operation or are capable of NLOS operation. Strictly speaking, however, the real distinction is not based on visible light, as the propagation behavior is dependent on radio waves of wavelength far greater than that of visible light.

A more accurate description is provided in IEEE Std 802.16.2-2004 [B26], which defines LOS as the condition in which the signal path is more than 60% clear of obstructions within the first Fresnel zone. The first Fresnel zone is defined as the region along the transmission path in which a reflected signal will be less than 90 degrees out of phase. Signals with an absolute phase offset less than 90 degrees add constructively to the received signal, whereas signals with a higher offset actually have a cancelling effect on the total received signal. It is hence desirable to receive the reflected signals of low phase shift.

Figure 3–1 illustrates an obstacle in the transmission path. The first Fresnel zone includes, at each point along the transmission axis [the line connecting transmitter (Tx) and receiver (Rx)], the circular area centered on that axis with radius $R_{F1}=\sqrt{(\lambda \cdot D_T \cdot D_R)/(D_T + D_R)}$, where λ is the wavelength of the RF carrier. The resultant ellipsoid is represented in Figure 3–1 by the dotted ellipse. For a single obstacle, the link is hence said to be LOS only if $R_o > 0.6R_{F1}$, where R_o is the distance of the obstruction from the transmission axis. The dashed line in Figure 3–1 shows an example reflection with exactly 90 degrees of phase offset.

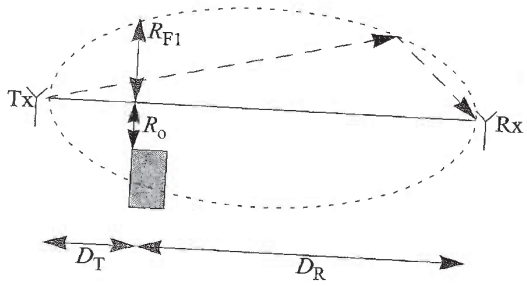


Figure 3–1: Fresnel zone

The definition of R_{F1} also makes immediately clear that for visible light, where λ is several hundred nanometers, the first Fresnel zone is many times smaller than for the spectrum relevant for IEEE Std 802.16 where λ is on the order of 1 cm to 1 m.

If multiple obstacles are present, the LOS condition is based on the minimum R_o/R_{F1}. Note that the ellipsoid is defined in three dimensions, and obstacles can exist not only to the side of the direct transmission path but also below and, in rare occurrences, above. Specifically, the curvature of the earth (which becomes more important as links get longer) can intrude into the 60% range if the antennas are placed insufficiently high.

Another manner of looking at propagation through reflections is by considering the effect not in the phase domain but in the time domain. This is most easily understood by the realization that reflections have a longer overall transmission path and hence, given the constant propagation speed of the signal, will arrive later and be more attenuated than the direct nonreflected path. This condition is termed *multipath propagation*. As a result, if symbols are transmitted sequentially without interruption, the receiver sees energy from one symbol smeared into the energy for the subsequent symbol.

In general, three methods can be used to mitigate the degrading effect of multipath on IEEE 802.16 networks. The first method is the use of very narrow beam antennas. This is suitable for LOS systems, since by this method the first Fresnel zone reflections will be only weakly stimulated. The second method is the use in the receiver of an equalizer that attempts intersymbol interference (ISI) cancellation by extracting and cancelling the energy of a symbol that has bled into subsequent symbols. In the third method, redundant samples are prepended to the transmitted waveform so that the receiver can discard these samples without loss of information. When sampling is cyclical in nature [as with the fast Fourier transform (FFT)], the prepended samples are termed the *cyclic prefix*. In that case, the receiver need not strictly discard the prefix, but may start its sampling already within the prefix if its duration exceeds the multipath duration. As described in later chapters, the cyclic prefix method is used in the OFDM and OFDMA PHYs of IEEE Std 802.16.

By use of methods like the latter two above, a system may be able to mitigate the effects of multipath propagation and communicate successfully even in NLOS environments. Such a system may be termed an *NLOS system*.

Multipath

When a radio signal is sent from a transmitter to a receiver, depending on the LOS and the reflecting objects, the signal may analytically split into multiple components with different delay, amplitude, and phase. These multiple components of the same signal may have positive or negative impacts on the received signal at the receiver, depending on the relative phases. In general, destructive interference is known as *multipath fading*, also termed *small-scale fading*. This is discussed in the next section.

Apart from the number of potential reflectors, the locations of the transmitter and receiver, and the distance between transmitter and receiver, it is important also to consider the impact of the orientation and movement of the communicating devices on multipath fading. Multipath causes the time offsets at which signals with substantial amplitude are received to lengthen. The energy of a signal transmitted with duration T_b is received with a duration longer than T_b. If the multipath duration is large compared to the symbol duration, the energy of a transmitted symbol will be received during a large multiple of the duration T_b and hence impact many subsequent symbols in a continuous symbol stream. This effect is termed *intersymbol interference* (ISI). A number of methods, know generally as *equalization*, exist to extract the energy of a transmitted symbol out of a received signal, but all of these methods get substantially more complex as the ratio of the multipath duration to the signal duration increases.

A related concept is *delay spread*. The delay spread defines the differences in the arrival time of signals that took different paths from the transmitter to reach the receiver. A good receiver has to be able to tolerate a certain delay spread. The delay spread usually grows with propagation distance (because the radii of the Fresnel zones increase as a function of distance) and, therefore, become significant in the design of outdoor wireless systems, where the distance between transmitter and receiver is typically long and the chances of multipath high as well. As the delay spread approaches the symbol duration, the probability of bit errors increases. The effect of multipath cannot be minimized by simply increasing the transmit power.

The problem of delay spread is minimized in low-frequency (below 11 GHz) IEEE 802.16 systems by either using multicarrier modulation (WirelessMAN-OFDM and WirelessMAN-OFDMA) or using equalizers (WirelessMAN-SCa).

Fading

Fading, which is the reduction of the received signal strength, is generally divided into three categories; large-, medium-, and small-scale. Large-scale fading, which is deterministic, is better known as *path loss* and depends on the separation between the transmitter and receiver. Medium-scale fading is also

known as *shadowing*. In general, it is the result of obstructions in the transmission path, such as rain and foliage.

Small-scale fading is in general the result of movement on transmission paths with multipath. Movement causes a shift in the signal's frequency as observed by the receiver, the so-called *Doppler frequency*, and causes rapid variations of the phases of the multipath components, the addition of which causes the rapid fading. The Doppler shift f_d itself can be calculated as $f_d = v \cdot \cos(\theta) \cdot f_c / c$, where v is the mobile's speed, θ the angle of movement relative to the direct line between transmitter and receiver, f_c the carrier frequency of the system, and c the speed of light. For example, for a mobile device traveling at 100 Km/h in a system operating at 2 GHz, the Doppler shift will range from –185 Hz to 185 Hz.

Small-scale fading is often characterized as nonfrequency selective (flat) or frequency selective. A signal is said to be flat faded if the coherence bandwidth of the channel exceeds the signal bandwidth; in other words, the fading of every frequency component is strongly correlated. A more detailed characterization of the above cursory description of fading can be found in [B49].

When a system is installed, its link budget is generally configured according to the large-scale fading with the addition of a margin for the cumulative effect of shadowing and small-scale fading.

ANTENNAS

Antennas are the physical elements responsible for converting the modulated signals into electromagnetic waves at the transmitter and performing the reverse operation at the receiver. The performance of an antenna constructed of usual materials is the same for both transmission and reception due to the fundamental property of reciprocity. A more detailed discussion of antenna technologies can be found in [B4], [B27], and [B47]. In this section, we discuss some of the basic antenna concepts.

Antenna parameters

The characteristics of the antennas used for wireless applications vary based on specific system requirements, such as deployment and coverage. The following is the commonly used set of parameters to characterize antennas:

- ***Gain:*** The gain of the antenna is the power output in a specific direction compared to that of a hypothetical isotropic antenna. The antenna gain is typically measured in dBi, the logarithmic scale of decibels relative to the isotropic antenna.
- ***Radiation pattern:*** The radiation pattern of an antenna is the graphical depiction of the antenna gain as a function of the elevation and azimuth angles. The radiation patterns for the horizontal and vertical planes are typically drawn on a polar graph as a function of the azimuth and elevation angles, respectively. A rectangular graph, instead of a polar graph, is sometimes used to illustrate the radiation patterns of narrow beam antennas.
- ***Beam width:*** The directionality of an antenna is indicated by its beam width. The beam width is measured in degrees between the points on the radiation pattern of the antenna at which the radiated power density is exactly half the maximum.
- ***Polarization:*** The polarization of an antenna is determined by the orientation of the electrical field component of the electromagnetic waves radiated by the antenna. Antennas usually provide linear, circular, or elliptical polarization. The linear polarization can be further classified into horizontal and vertical polarizations.
- ***Bandwidth:*** When an antenna is designed to operate in multiple frequency bands, it is important that it maintain uniform impedance across all those frequencies. Antenna impedance matching plays an important role in maximum power transfer between transmitter or receiver circuitry and free space. The band of frequencies in which an antenna is designed to operate is known as the *antenna bandwidth*.

Apart from the system requirements, regulatory rules may also limit the type of antennas that can be used. For example, many regulatory domains have specific limits on antenna gain and/or beam width. Some regulatory domains

may limit the effective isotropic radiated power (EIRP), which is a sum of transmit power and antenna gain, minus any loss in the transmit path (all in decibels).

Directional and sectorized antennas

The antennas used in PtP, PMP, and mesh systems usually have different beam widths. Omnidirectional antennas are rarely used in fixed outdoor PtP systems, given the range requirements of the outdoor systems and the need to minimize interference. Directional antennas have important characteristics that are of significance to BWA; i.e., they increase the amount of power transmitted in a specific direction and reject the signals coming from other directions on reception. For PtP systems, highly directional antennas are used, as the goal is to establish communication only between two specific and fixed end points.

PMP systems typically deploy sectorized antennas, with beam width greater than 15 degrees, at the BS to improve frequency reuse in cellular systems. These minimize the interference across cells, as compared to omnidirectional antennas. Sectorized systems use more than one sector to cover all 360 degrees, if necessary. As many as six sectors are used in many practical systems.

Even in the case of PMP, the UL antenna may, like the PtP case, be highly directional and pointed at the BS. The tradeoff for such a high-gain antenna is the need for a LOS and an expensive professional outdoor installation that may call for careful alignment. Consumer-based applications, particularly at the lower frequencies, may call for self-installation and, therefore, require a much different (perhaps omnidirectional) SS antenna.

Diversity

Diversity is a technique used in wireless communication to mitigate the effects of multipath fading and to improve performance by exploiting the fading characteristics of wireless transmissions. For example, the *temporal* and *spectral diversity* schemes make use of the time-selective and frequency-selective fading characteristics of channels, respectively. When neither of

these schemes is possible or desired, *spatial diversity* schemes may be used. In spatial diversity, more than one antenna is used in the transmitter and/or the receiver to make use of the spatially independent fading characteristics of the channel. Since spatial diversity requires more than one antenna, it is sometimes referred to as *antenna diversity*. Depending on the use of multiple antennas at the transmitter or receiver, it may be called *transmit antenna* or *receive antenna diversity*.

Multiple diversity schemes may also be combined to provide improved performance. For example, spread spectrum and OFDM schemes essentially make use of spectral diversity of the frequency selective characteristics of the channel. Spatial diversity schemes can also be used in a spread spectrum or OFDM system. The following subsections briefly describe temporal, frequency, space, angle, and polarization diversity.

Temporal diversity

In temporal diversity method, the signal is repeated in time, where the time interval between repetitions is sufficiently large to have different channel characteristics. Therefore, the received signals are expected to have independent fading.

Frequency diversity

As the name implies, frequency diversity involves the transmission of the signals in frequencies spaced sufficiently far apart to cause independent fading. Frequency diversity is typically exploited only in multicarrier systems, since the use of more than one frequency is otherwise not practical. In IEEE 802.16 multicarrier PHYs, frequency diversity is exploited through the use of interleavers and the mapping of subchannels to physical carriers. Frequency diversity may be implemented with one antenna.

Spatial diversity

Spatial diversity schemes can be classified into *diversity combining* and *switched diversity* schemes. With diversity combining, the signals received from more than one antenna are combined to reduce the effect of amplitude

fading due to multipath. The receiver typically uses switched (selection) combining, maximum ratio combining (MRC), or equal gain combining (EGC). In switched combining, antennas are placed sufficiently apart that the fading is independent on each receiver antenna. The combining process takes the best signal from the receive antennas. MRC requires the estimation of the channels and weights the signals received at each receive antenna, per the estimation, in order to maximize the signal to noise ratio (SNR). EGC is very similar to MRC except that the weights are independent of the channel estimates. Instead, the signals from each receive antenna are multiplied by the same weight. Since there is no channel estimation in EGC, the SNR gain of EGC is lower than that of MRC.

Diversity combining is sometimes impractical due to its complexity and cost. The simpler switched diversity schemes may be used to switch between one of the available antennas based on known characteristics or some dynamic measurements. The switched diversity schemes can be implemented at either the transmitter or the receiver. The popular IEEE 802.11 WLAN devices often use transmit diversity with two antennas. The receive antennas can also be chosen similarly.

A variation of receiver diversity known as *fast-receive diversity* is sometimes used to adaptively choose the antenna with the best signal for reception. However, fast diversity schemes cannot be supported if the preamble is too short, as there may be insufficient time to detect and measure signals and adaptively switch between antennas. The goal of switched diversity schemes is to mitigate the effects of multipath, not to increase the gain.

All the spatial diversity techniques require at least two antennas. The diversity combining techniques use at least two antennas simultaneously. However, the switched diversity techniques use but one transmit or receive antenna at any given time.

Polarization diversity

Polarization diversity makes use of the orthogonality of polarized transmissions. Since the fadings for horizontally and vertically polarized signals are often uncorrelated due to the phase and amplitude differences, polarization diversity can be implemented with a single antenna that is

capable of receiving both vertically and horizontally polarized signals. In practice, polarization diversity is often implemented with two antennas, one horizontally polarized and the other vertically polarized. When space is limited, polarization diversity can be helpful since the antennas can be placed together. It should also be noted that, in a highly reflective environment, the polarization of the transmitted signals may be lost; therefore, polarization diversity may provide only minimal gain.

Angle diversity

With angle diversity, the pointing angles of the antennas are different. This technique makes use of the uncorrelated signals arriving at different angles. Angle diversity requires at least two antennas. As with polarization diversity, angle diversity is useful when the location of the antennas are space constrained.

Multiple antenna systems

Multiples antennas may be used on a single wireless device for a variety of reasons. The most common application of multiple antennas is to support one of the diversity schemes described in the previous section. Adaptive antenna systems (AAS) and multiple-input, multiple-output (MIMO) systems also exploit the properties of multiple antennas, as described below.

Adaptive antenna systems (AAS)

Adaptive antennas can adaptively adjust to meet specific performance goals, such as improved link margin or improved signal to interference ratio (S/I). Although adaptive antennas can be implemented in many different ways, they are classified into one of the following three categories:

- ***Switched beam:*** This is the simplest form of adaptive antenna design, where the antenna pattern of each antenna is fixed. The best antenna for communicating with a specific remote wireless device is chosen adaptively based on a variety of factors such as past measurements from each antenna. The antenna selection concept of switched beam architecture is very similar to that of switched antenna diversity.

- ***Beam steering:*** Unlike the fixed beams of switched beam design, the beam steering approach is used to adaptively steer the beams toward the remote wireless device to maximize the gain. The objective is to maximize the gain from and to every wireless device.
- ***Optimum combining:*** The optimum combining method uses a linear spatial filter to adaptively adjust the antenna, with a periodically adjusted reference signal, through feedback. The linear spatial filter is used to suppress the noise and interference, while separating the desired signal. Optimum combining should not be confused with the aforementioned MRC, as MRC is not spatially discriminatory and is optimal only when the noise is white (i.e., uncorrelated at the different antennas) and there is no interference.

The gain and S/I achieved by switched beam adaptive antennas are limited by the pattern, the angle, the interference, and the LOS conditions of the area covered by the fixed antenna. The complexity of the beam steering and optimum combining approaches also depends on whether the wireless devices are fixed or mobile. For the beam steering approach, maximizing the gain for each wireless device requires that the power be controlled for every beam. While the path loss may not change significantly for a fixed wireless device, a mobile device may require dynamic power adjustments to the beams.

The optimum combining approach depends on a feedback system and requires dynamic update of filter parameters to account for fading. Since the fading rate of mobile devices can be much higher, the complexity of processing the updates in mobile networks is typically higher than that of fixed networks.

Multiple-input, multiple-output (MIMO)

MIMO systems make use of multiple antennas at both ends of the link. In a MIMO system, potentially uncorrelated signals from different antennas can be used to maximize capacity. Given this property, the gain from MIMO in LOS or PtP links will be minimal, as the signals in such environments will be highly correlated. However, MIMO offers many benefits in a typical NLOS environment, where the multipath effect leads to highly uncorrelated signals.

The main difference between adaptive antennas and MIMO is that the adaptive antenna schemes achieve capacity gain by suppressing interference through directional discrimination, allowing simultaneous transmissions to different terminals. MIMO, on the other hand, achieves the same by exploiting uncorrelated signal paths to, in effect, transmit multiple data streams simultaneously to a single terminal.

Impact of antenna technologies on protocol design

In any wireless system, the MAC protocol design must consider the antenna technology. While some MAC protocols may work equally well on a variety of antenna types, some protocols are optimized for or constrained by specific antenna types. For example, if the protocol depends on "listen before talk" schemes such as CSMA, any directionality may work against the efficiency of the protocol because some receivers are inherently unable to detect some transmitters and cannot, therefore, sense ongoing transmissions.

On the other hand, omnidirectional antennas support a broadcast-type network that can be used to minimize the control traffic overhead. The lack of broadcast capability can increase the overhead in multiple antenna systems such as the AAS. Therefore, the MAC protocols designed for multiple antenna systems need to optimize the control traffic overhead without sacrificing other benefits of such a system. There may also be some implementation constraints in multisector systems, where the MAC transmissions and receptions may need to be coordinated across sectors, depending on the frequency plan.

Antenna design for fixed and mobile devices

The choice of antenna technology also varies depending on whether a system or device is fixed or mobile. Since a typical mobile system can include both fixed devices such as BSs and mobile devices such as handsets, the fixed and mobile devices in the same system may use different types of antennas. For example, most mobile devices are too small for directional antennas; even if they were not, orienting such antennas would be problematic. Therefore, most mobile devices use omnidirectional antennas, even if the BSs use directional or sectorized antennas.

The more complex adaptive antennas are also difficult to implement on a mobile handset using current technologies. Similarly, the support for antenna diversity with sufficiently spaced antennas is limited, depending on the frequency of operation, due to the size. Therefore, some of these adaptive antenna and diversity methods are used only in a fixed system or the fixed part of a cellular system. Some types of diversity, such as polarization diversity, are also unsuitable for a mobile device since the polarization is difficult to control.

PHYSICAL LAYER (PHY)

Forward error correction (FEC)

One way to deal with errors in wireless transmissions is to automatically detect and correct symbol or bit errors rather than relying on the detection and retransmission of packet errors [e.g., using checksum or cyclic redundancy check (CRC)]. The most commonly used technique for error correction is called *forward error correction* (FEC), which is capable of detecting and correcting some errors upon reception by adding redundancy to the transmitted signals.

Wireless protocol design must consider the overhead and cost associated with PHY-level error control schemes, such as FEC, and MAC-level error control schemes, such as automatic repeat request (ARQ). For example, while FEC may be able to correct certain types of errors quickly with high probability, the addition of FEC to every transmitted block reduces the efficiency of the channel and could increase the delay of good protocol data units (PDUs) due to longer interleaving. On the other hand, MAC-level ARQ can increase the delays in a channel with high error rate, where the detection, communication, and retransmission can consume a significant portion of the bandwidth. Many systems support hybrid techniques, where a combination of FEC and ARQ parameters may be adjusted, often adaptively, to offer the desired performance.

Single carrier and multicarrier

As described above, the delay spread and fading caused by multipath are challenging problems that any practical outdoor, or outdoor-to-indoor, wireless system must solve. While equalizers can be used to offset the effect of multipath in a single-carrier protocol, the complexity of time domain equalization and associated channel estimation increases with higher data rates and time-varying channels. In most BWA cases, the delay-spread–induced ISI is significant only for high signal rate, typically above 1 Msymbol/s (i.e., symbol durations shorter than 1 μs). If the signalling rate can be reduced by increasing the symbol duration, the effect of ISI on the channel can also be reduced. However, high signal rates are desirable for high data rates. In multicarrier modulation schemes, the available frequency band of a given channel is divided into a large set of (typically, orthogonal) subcarriers, each of which is modulated separately. This leads to a much smaller signal rate per subcarrier compared to single-carrier systems. The data rate achievable over these simultaneously transmitted subcarriers is almost equivalent to the data rate of a high-rate single-carrier transmission, with the difference stemming from the cyclic prefix overhead. Multicarrier modulation schemes, such as OFDM (see [B55]), do come with other drawbacks, including the higher sensitivity to frequency offset (the equivalent of timing sensitivity in other systems) and phase noise. They also require a somewhat larger power amplifier backoff due to higher peak to average power ratio (PAPR).

Instead of time-domain equalization, a single-carrier system may also employ frequency-domain equalization. In this case, the receiver contains an inverse fast Fourier transform (IFFT) and FFT pair, with channel estimation executed between the transforms. The result is some characteristics similar to those of an OFDM transceiver, including similar complexity.

DUPLEXING, MULTIPLEXING, AND MULTIPLE ACCESS

The following subsections provide a brief overview of some of the key duplexing, multiplexing and multiple access concepts. For a detailed discussion on these and related concepts, the reader is referred to [B4] and [B7].

Duplexing

Duplexing defines how bidirectional communication is achieved between two devices or between a BS and a set of client devices in a PMP system. The communicating devices themselves may be capable of half-duplex (transmit or receive but not both simultaneously) or full-duplex (transmit and receive simultaneously) operation. There are two types of duplexing: *time division duplexing* (TDD) and *frequency division duplexing* (FDD). Figure 3–2 illustrates TDD communication in a PMP system, with DL and UL on the same frequency channel. TDD is a half-duplex method by definition; i.e., when one device transmits, the other receives, and vice versa.

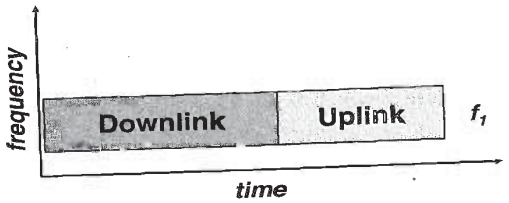


Figure 3–2: TDD illustration

In FDD, the devices use different frequencies to transmit and receive. Therefore, it is possible to support full-duplex communication in an FDD system. Figure 3–3 shows full-duplex operation, where the transmission and reception overlap in time. However, half-duplex frequency division duplexing (H-FDD) operation, as shown in Figure 3–4, is also possible, as is a mix of full-duplex FDD and H-FDD devices in the same system. An H-FDD device is usually cheaper to build as it requires only one transmitter and one receiver. The H-FDD operation is also referred to as *frequency shift division duplexing* (FSDD). Alternate expansions of the abbreviation FSDD include *frequency switched division duplexing* and *frequency simplex division duplexing*.

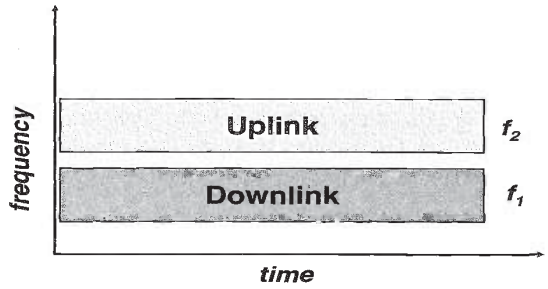


Figure 3–3: Full-duplex FDD illustration

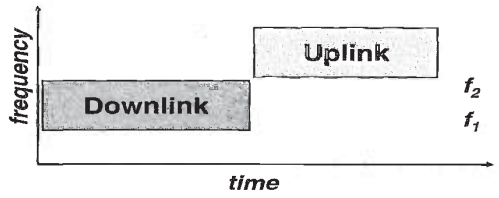


Figure 3–4: H-FDD illustration

In TDD, the time allocated for each direction (i.e., UL or DL) may be fixed or adaptive. Adaptive time division duplexing (ATDD) is a variation of TDD in which the UL/DL ratio can be dynamically adjusted per frame, based on traffic conditions and other system parameters. ATDD is of great interest to broadband wireless systems, due to its flexibility in supporting asymmetric traffic and efficient use of the spectrum. However, in cellular deployments with multiple ATDD systems, if frequency reuse is not properly employed, ATTD may increase the level of interference. In some practical multisector systems, the ATDD boundaries may need to be coordinated across sectors, even if the sectors operate on different frequency channels due to adjacent channel interference issues.

FDD may require more spectrum than a comparable TDD system, increasing the potential for interference, especially in license-exempt spectrum. The frequency channels used in both directions also need to be sufficiently separated to avoid interference. Moreover, in access networks, spectrum may be wasted if FDD uses symmetric channels in both directions, as the traffic load may be asymmetrical. On the other hand, the QoS scheduling and channel access functions may be less complex in FDD compared to TDD systems.

The duplexing method is independent of the multiple access and multiplexing methods. FDD is commonly used in licensed voice-centric networks such as cellular telephone systems, where most of the traffic has historically been symmetric.

Multiplexing

Multiplexing refers to a mechanism in which a single device transmits to multiple devices on a single channel. In time division multiplexing (TDM), the transmitting device divides the time domain into multiple slots to communicate with multiple devices. Needless to say, TDM requires only one frequency, but the available time must be divided among the potential receivers. In frequency division multiplexing (FDM), the transmitting device uses different frequencies to communicate with multiple devices. Any wireless protocol design has to balance this flexibility against the spectrum usage requirements.

Centralized and distributed multiple access schemes

Multiple access refers to the way that multiple devices access the medium, regardless of whether the communication is many-to-one or many-to-many. Distributed access methods such as CSMA require no centralized coordinator. While this may be simpler to implement, the spectrum efficiency of this method is low; and the potential for collisions affects the overall system throughput and scalability, causing a potentially dramatic decline in efficiency as the system load increases. Collisions also introduce nondeterminism into the system, thereby making QoS guarantees difficult to achieve.

Time division multiple access (TDMA) and frequency division multiple access (FDMA)

In TDMA, multiple devices use their own predetermined time slots to transmit. The number and location of the time slots may be statically determined or dynamically assigned. If dynamically assigned, some form of coordination (centralized or distributed) is necessary. It is possible to provide a deterministic access to the medium using a TDMA scheme. For example, IEEE Std 802.16 relies on the BS to allocate time slots for UL transmissions. The BS also schedules DL transmissions in a similar fashion, using TDMA bursts addressed at specific groups of SSs. Some other PMP systems use continuous TDM in the DL, with the information for each SS multiplexed onto a single stream of data directed to all SSs within the sector. The continuous TDM stream does not require preambles for resynchronization between bursts. The downside, however, is that it does not allow the flexibility of adaptively adjusting the coding and modulation to specific SSs.

FDMA is similar, with multiple devices using different frequencies to access the medium. However, the amount of spectrum needed and the co-channel and adjacent channel interference can decrease the efficiency of the FDMA systems. While traditional TDMA systems used fixed allocations to the existing devices, which may result in wastage of bandwidth if not used, it is possible to allocate bandwidth dynamically in a TDMA system, where the slot allocations are changed and communicated dynamically. Both TDMA and FDMA can be used with FDD or TDD. Hybrid or combination methods are also possible.

Orthogonal frequency division multiple access (OFDMA)

As described earlier in this chapter, OFDM is a multicarrier modulation method, but not a multiple access method. OFDM may be used with many multiple access methods, including CSMA, TDMA, and FDMA. The IEEE 802.11a and IEEE 802.11g amendments use an OFDM PHY with CSMA, and the IEEE 802.16 OFDM PHY option is used with TDMA. However, OFDM's use of multiple subcarriers can be used as the basis of a multiple access method. The extension of OFDM as a multiple access method is called *orthogonal frequency division multiple access* (OFDMA), in which a

subset of the mutually orthogonal subcarriers may be allocated to a specific wireless device for access to the channel. The concept is a sophisticated form of FDMA, with the orthogonality of the subcarriers making it easier to separate them without the need for guard bands.

Many variations of OFDMA exist, distinguished primarily on the basis of how the subcarriers are allocated to users. For example, the subcarriers may be allocated in a frequency-hopping pattern. They may also be allocated dynamically by a central controller. The difference between OFDM and OFDMA is illustrated in Figure 3–5. While all the subcarriers will be used by a single user in OFDM, the subcarriers may be shared by a number of users in OFDMA. The number of subcarriers assigned to each user may vary based on system parameters and demand.

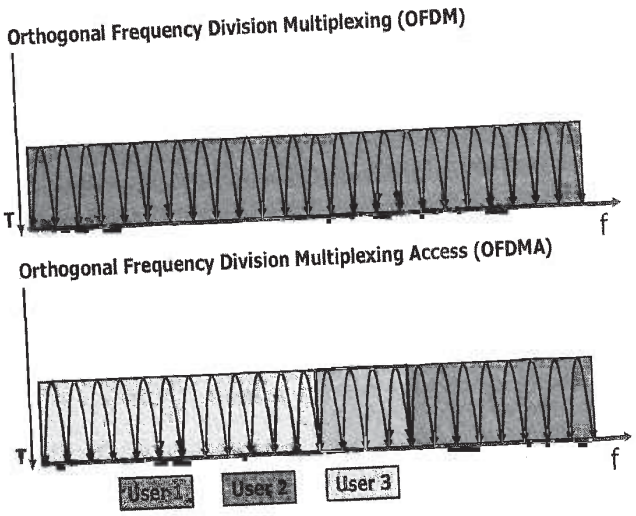


Figure 3–5: OFDM and OFDMA

Code division multiple access (CDMA)

In CDMA, each user is assigned a unique code used to encode the transmissions on a channel. Instead of transmitting a single data waveform, the transmitter sends the time-domain convolution of the data waveform with the code so that the transmission rate (often referred to as the *chip rate*) is many times higher than the data rate. The high transmit rate, and hence short duration of each chip (code element), results in a spreading of the signal in the frequency domain. This is why these codes are often referred to as *spreading codes*.

At the receiver, the convolution of the signal with the known sequence allows the retrieval of the original data symbol while spreading out any narrowband interference in the channel. CDMA is a popular choice for military systems for the following reasons:

1. The requirement that transmission code be known in order to receive the signal
2. The spreading of narrowband interference
3. The fact that the spreading of the signal in the frequency domain can lower the energy per hertz to well below the thermal noise floor so that its existence is not obvious to third parties

The design of CDMA codes is based on minimizing the autocorrelation of the code and the cross-correlation of the code with all other used codes. The better this design, the easier it is for the receiver to separate the simultaneously transmitted codes, and the more robust the system is to power variations between terminals.

Since multiple codes can be used simultaneously on the same channel, CDMA can be a bandwidth-efficient solution that requires no coordination or synchronization.

In the OFDMA PHY described in Chapter 12, the CDMA concept is also used for contention-based access. In that case, however, the spreading sequence is not transmitted in time, but each chip is modulated on a separate subcarrier before the IFFT.

DATA UNITS

Data units are the basic units exchanged between different layers of the protocol stack. The service data unit (SDU) and the protocol data unit (PDU) are the two fundamental data units used here. The SDU is the data unit exchanged between two adjacent protocol layers of the same device. The service access point (SAP) defines the interface between two adjacent protocol layers, where the services of the lower layers are available to the higher layers. The SDUs are exchanged between two adjacent protocol layers through the SAP, as the lower layers provide the services of accepting the SDU from the higher layer and delivering the SDU to a higher layer. The PDU is the data unit exchanged between the peer entities of the same protocol layer. When a PDU is transferred to a lower layer for transmission, it becomes the SDU of the lower layer.

An SDU received from a higher layer may go through a variety of transformations to become a PDU for transmission to a peer layer. When the MAC or PHY receives an SDU, it may add additional headers and control information to reliably transmit the PDU to the peer. In addition, SDUs may go through fragmentation, in which more than one PDU may be constructed from a single SDU. Similarly, fragments of multiple SDUs or multiple full SDUs may be packed into a single PDU. The only basic requirement is that the inverse transformation must be performed by the same peer layer on the receiving end before delivering the PDU to a higher layer. Figure 3–6 shows the concept of PDUs, SDUs, and SAPs.

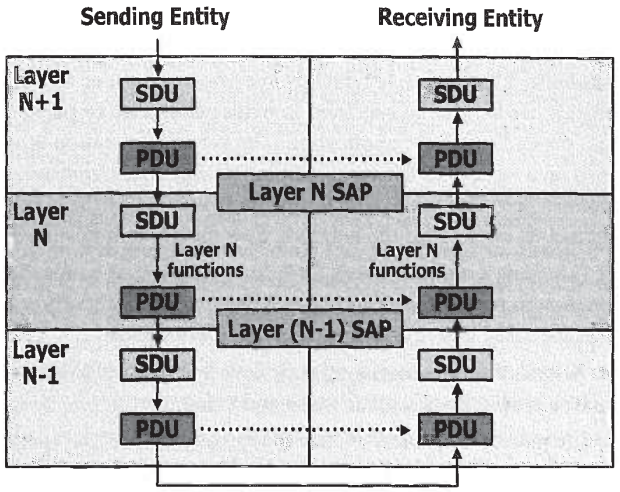


Figure 3–6: PDUs, SDUs and SAPs

QUALITY OF SERVICE (QoS)

A BWA system supporting multiple services must properly account for the fact that different services require different performance levels in order to maintain the appropriate QoS. A network that supports differentiated QoS can provide services at specified performance levels using a set of qualitative and quantitative measures.

Applications such as voice, video, and data transfer demand various QoS requirements. The QoS requirements of the application traffic can be specified and enforced either as a set of priorities or as a set of parameters. In the prioritized QoS model, the traffic is treated by the network based on relative priority at any instant. However, in a parameterized QoS model, the network attempts to guarantee a set of QoS parameters for the traffic, in conformance with a preagreed or configured service.

The most commonly used QoS metrics in the parameterized QoS model are throughput, delay, jitter, and loss:

- ***Throughput:*** Throughput is typically specified in bits per second, although it can be also be specified in bytes per second or packets per second, depending on the application. The actual application throughput depends on a variety of factors, including packet size, overhead, and retransmissions. The achievable throughput is bound by the maximum and instantaneous capacity of the network, also specified in bits per second.
- ***Delay:*** The delay, also known as *latency*, is the time taken for the information to travel from a source to a destination and vice versa. The delay is specified in units of time, such as milliseconds.
- ***Jitter:*** Jitter is the variation in delay. It is very important to many interactive applications such as voice and video.
- ***Loss:*** The packet loss in any system is typically defined as a percentage. Data applications can tolerate some delay and jitter, but they cannot readily tolerate packet loss. If a data packet is lost, it must be retransmitted by some layer (e.g., network or application) for the application to work correctly. On the other hand, real-time applications such as voice and video can tolerate a small percentage of loss. Retransmission can be used to compensate for packet loss at the cost of increased latency and reduced available capacity.

The simplest approach to QoS is static provisioning of services and reservation of resources. This is how circuit-switched networks guarantee QoS, where a circuit is established by reservation of specific resources for the duration of a connection or call. However, this is not a practical or efficient solution in packet networks, since not all users are expected to be simultaneously active and not all active users are expected to use a fixed percent of the network resources. The challenge in such networks is to support adaptive QoS techniques that can utilize the available resources more efficiently.

Therefore, in order to fully support QoS at the network layer, the following four functions are required:

- ***Admission control and provisioning:*** Irrespective of how a network solves the QoS problem, networks are constrained by the maximum amount of traffic and services they can support at any particular instant. Therefore, it is important to regulate which users can access the network and how they do so. The control must be dynamic, based on the network load and other parameters. This process is called *admission control* and, without it, networks may become oversubscribed, defeating the purpose of QoS guarantees. *Provisioning* is a related concept. Networks may be statically or dynamically provisioned for specific traffic situations. Static provisioning is often used in conjunction with admission control to over-provision the network and then manage admissions dynamically. While a variety of algorithms can be implemented to perform admission control, the underlying network must provide certain functions to support intelligent admission control functions. A simple admission control algorithm may merely admit or deny new requests based on the number of previously admitted services. However, a more sophisticated algorithm may try to adjust the parameters of admitted services to accommodate the new request. In order to support such admission control functions, a network must provide basic support for modifying admitted service parameters gracefully without affecting the performance or increasing the overhead. Except for such support functions at the network, admission control is primarily an independent function. Therefore, the specification of the admission control algorithm is outside the scope of an air interface standard. Vendors are free to implement specific admission control algorithms to differentiate their products.
- ***Traffic classification:*** In order for the network to provide differential treatment based on QoS requirements, it must be able to classify the incoming packets. The classification process may use some basic header information from a packet to determine how it should be treated. Some QoS-aware networks may also support specific identifying information in each packet to carry the classification or priority information so that explicit traffic classifications at every step (e.g., every hop or every layer of a protocol stack) can be avoided. As we will see in later chapters, the IEEE 802.16 MAC is connection-based and can be used to associate a set of QoS parameters after classification or to map preclassified packets from another protocol to a specific connection.

- ***Shaping and policing:*** The admission control manages the admission of a specific traffic stream or flow into the network. However, the admitted traffic still needs to be regulated dynamically to make sure it conforms to the agreed-upon service specification or profile. Shaping and policing are the two commonly used techniques to monitor and regulate incoming traffic for conformance. *Policing* merely discards or reclassifies the nonconforming packets to a lower priority or class. *Shaping*, on the other hand, may queue the nonconforming packets and delay the release of the packet into the network. The algorithms for shaping and policing are not typically specified in a networking standard. However, a standard may support the necessary signaling and management functions to specify shaping and policing parameters.
- ***Traffic scheduling:*** The traffic scheduler is responsible for determining the transmission order of the packets based on the QoS requirements, when fluctuations in the incoming rate results in queueing of the traffic. Therefore, the scheduler is an important component of any QoS architecture. A network without QoS support treats all types of application traffic alike in a first-in, first-out (FIFO) fashion. A QoS scheduler, on the other hand, has to ensure that the traffic is serviced based on the QoS requirements and not based solely on the order of arrival. However, fairness is still an important criterion for any QoS scheduler, and the higher priority traffic should not completely starve the lower priority traffic. As we will discuss later, wireless QoS scheduling has some important differences compared to wired QoS scheduling. The networking standards may define specific scheduling behaviors and support functions for QoS signaling and enforcement. However, a specific scheduling algorithm is typically not specified by a standard. IEEE Std 802.16 is no exception and leaves the scheduling algorithm as a vendor differentiator.

Per-flow and per-class QoS

QoS can also be enforced at varying levels of granularity. For example, each flow (e.g., every voice conversation) can be classified into a separate stream and handled separately by a network. Alternatively, each class of traffic (e.g., all voice traffic or all video traffic) can be classified into a separate stream.

For a complete discussion on per-class QoS, per-flow QoS, or other ways of differentiating QoS, see [B33] and [B42].

As a general rule, it is practical and sufficient to provide coarse-grain QoS only in core aggregated networks or backbone networks. It would be impractical to maintain state for a large number of flows in a network that carries traffic from thousands of sources. However, fine-grain QoS may be practical and is often necessary at the edge of the network, where the number of unique flows can be smaller and the need for QoS differentiation can be much greater, given the limited availability of resources and other constraints. While some network architectures may support only per-class QoS, others may support both per-class and per-flow QoS.

Wireless QoS

Enforcing QoS can be simply viewed as efficiently reordering transmissions to guarantee throughput, delay, and jitter requirements of all traffic. However, QoS enforcement has fundamentally different challenges in wired and wireless systems. In wireless networks, QoS is not just the reordering of packets. While enforcing QoS in any network in an efficient manner with conflicting requirements can be challenging, most wired networks fall under the class of fixed-capacity networks, where the maximum available capacity does not change over time. Also, the probability of errors in wired networks such as Ethernet or ATM is extremely low. Therefore, a QoS enforcement algorithm in a wired network is primarily responsible for optimal reordering of transmissions. This also simplifies the admission control process, as the instantaneous capacity is fixed and is always known.

On the other hand, wireless networks are varying-capacity networks, where the instantaneous capacity varies based on the conditions of the RF environment. This brings in additional constraints so that scheduling of bytes or packets to enforce QoS may not be sufficient or fair. For example, the airtime required to transmit a byte varies with the modulation and coding scheme. The changing RF conditions often force the modulation and coding rate to be changed dynamically, introducing an additional constraint for the QoS scheduler.

Another important dimension in wireless QoS is the ability of a scheduler to deal with high error rate. For example, if a transmission failed to reach the receiver due to large-scale fading, it may not be advisable to retransmit the data immediately, as transmission is likely to fail again. A high retransmission rate for a specific receiver not only increases the average delay for that receiver but also affects the QoS of others in the system. The QoS scheduler has a challenging task of maintaining fairness in such a system, balancing throughput and the use of over-the-air resources. Dynamic fragmentation is one way to address fairness, fragmenting SDUs dynamically to fairly divide the airtime.

If the wireless system supports ATDD with centralized scheduling (i.e., the BS is responsible for QoS in both directions), the scheduler also has to decide on a frame-by-frame basis how to share the channel between the UL and DL.

MEDIUM ACCESS CONTROL LAYER (MAC)

The MAC function is responsible for controlling access to the medium. MAC protocols may use distributed multiple access techniques, or centrally coordinate the access for all devices, or use a combination of the two, as discussed earlier.

Apart from multiple access, MAC is also responsible for basic functions such as data encapsulation, fragmentation, and adaptive modulation support. A wireless MAC must also define how MAC protocol data unit (MPDU) errors are detected and how, if necessary, the faulty MPDUs are retransmitted. QoS and security are two other functions that must be supported in a wireless MAC, given the limited bandwidth and reduced level of physical security in comparison to wired networks. The MAC, along with the LLC, is Layer 2 of the OSI reference model.

Fragmentation and packing

The MAC is responsible for encapsulating higher layer packets or SDUs in its own PDU format and delivering it to the peer MAC on another device. One simple way to encapsulate a higher layer SDU is to add necessary information and deliver the whole MPDU to the peer MAC. However, there are many

reasons for the MAC to fragment a single SDU into multiple fragments before transmission, reassembling those fragments at the other end before delivering the resultant PDU to the higher layers. Fragmentation is necessary to efficiently support higher layer protocols with variable-size SDUs, such as Internet Protocol (IP), if the underlying MAC and PHY are based on a framed architecture (i.e., the frame size and, therefore, the available airtime per frame is fixed for a specific deployment). Some rationale and benefits of fragmentation include the following:

- ***Probability of successful delivery***: The probability of successful delivery may depend on the size of the MPDU, especially in random access MAC or lower data rate transmission, where the chances of collision or other failure increases with the size of MPDUs.
- ***QoS***: There are many QoS-related reasons for fragmentation. A low-priority, but large, SDU may need to be fragmented so that it does not block a smaller, but higher priority, SDU. The so-called *head-of-line blocking problem* may be alleviated by fragmentation. Another QoS-related reason is to minimize the overall delay in an error-prone channel, reducing the probability of a larger SDU being retransmitted.
- ***Spectral efficiency***: Fragmentation in an error-free channel may increase the MAC overhead, thereby adversely affecting spectral efficiency. On the other hand, the cost of retransmitting a large lost MPDU is higher than the cost of retransmitting a smaller fragment of a large SDU. Another important aspect of efficiency is in a framed MAC, where the data have to fit within predefined frame boundaries. In order to fit MPDUs into the available frame time without wasting the airtime, it is important to dynamically fragment SDUs and fit the MPDUs into a frame.

While fragmentation is useful with large SDUs, many communication protocols have smaller SDUs to transport. An example is the widely used networking protocol, Transmission Control Protocol (TCP), in which 40-byte acknowledgments traverse the reverse direction of the data flow and, on average, one acknowledgment is sent for every two TCP packets. Voice application also uses a small payload size to keep the delay at a minimum. Since MAC encapsulation introduces some fixed overhead per packet, it is beneficial for the MAC to allow the packing of multiple higher layer payloads

into a single MAC payload, without affecting the delay or fairness of the network.

Automatic repeat request (ARQ)

Any wireless system must define a process for handling MPDU errors. The PHY may support error correction methods to minimize the number of errors. Nonetheless, additional error correction methods are necessary, especially in a wireless MAC, as bit errors are inevitable. The ARQ is a class of retransmission algorithms for supporting reliable delivery in the presence of errors. An ARQ algorithm must define the following:

- Error detection
- Feedback policy
- Retransmission strategy
- Retransmission unit

The detection of MPDU errors is typically handled by adding a checksum or a CRC to every MPDU transmitted. It is important to choose a CRC or checksum that has an extremely low probability of undetected errors and minimal overhead to the protocol. The FEC described in the PHY section earlier in this chapter may also be used as an error detection mechanism, when the FEC detects errors but cannot correct them.

An important part of the retransmission algorithm is determining which MPDUs have not reached the receiver. The design of the ARQ feedback policy needs to consider the overhead associated with the feedback and the potential delay introduced by a specific policy. The ARQ feedback may include positive and/or negative acknowledgments from the receiver, commonly referred to as *ACKs*. Acknowledgments may be sent for every MPDU or for a number of MPDUs in an aggregated form. If the acknowledgment is per-MPDU, then the choice is to send the acknowledgment either immediately following every MPDU or at a later time.

The baseline IEEE 802.11 WLAN MAC used an immediate per-MPDU acknowledgment for all directed MPDUs. While it is possible to make

retransmission decisions quickly in this type of acknowledgment strategy, the overhead can be high. A variation of this might be to send only negative acknowledgments instead of positive acknowledgments. If the error rate is low, this might reduce the overhead and keep the delay under control. Of course, the positive and negative acknowledgments can be combined to provide a complete feedback to the transmitter. Another way to minimize acknowledgment overhead is to use a bitmap to indicate positive and negative acknowledgments.

Once an error is detected and the missed MPDU information is communicated to the transmitter, a retransmission policy decides when and how MPDUs are retransmitted. The retransmission policies can be broadly classified into three types: *stop-and-wait*, *go-back-n* (GBN), and *selective repeat* (SR) (see [B6] and [B9]). In stop-and-wait retransmission strategy, the transmitter waits for an explicit acknowledgment before deciding to retransmit the same MPDU or to transmit the next MPDU to the same receiver. While this is a simple algorithm, the strategy can introduce higher delays even if the acknowledgment immediately follows the MPDU.

The GBN algorithm retransmits MPDUs starting from the missing MPDU, including MPDUs that were possibly received correctly. This algorithm is attractive from the perspective of the receiver buffering, as the receiver does not have to buffer out-of-order MPDUs. However, this algorithm makes inefficient use of the wireless medium, and it is possible for an MPDU that was successfully received to get retransmitted multiple times just because an earlier MPDU failed more than once.

The SR algorithm selectively retransmits only the MPDUs confirmed or assumed to be lost. While this algorithm makes efficient use of the medium, the receiver has to buffer out-of-order MPDUs so that they can be delivered to the higher layers in order. SR has been used in many wireless and mobile standards such as European Telecommunications Standards Institute (ETSI) Broadband Radio Access Networks [Technical Committee] (BRAN) HiperLAN/2 (see [B14] and [B36]), GPRS (see [B15] and [B41]), and some IMT-2000 specifications (see [B31]). Since wireless resources are scarce and SR provides higher channel efficiency, one of the many variations of SR has been the preferred algorithm for many wireless systems.

While ARQ can improve the reliability of wireless communication, a certain level of basic reliability is still necessary for a wireless communication to be practical. For example, it is debatable how much ARQ can help time-sensitive traffic such as voice. Depending on the additional delays introduced by a specific ARQ algorithm, ARQ may be unsuitable for voice channels or may be suitable in only limited cases.

Chapter 4 IEEE 802.16 architecture

Overview and key features

In this chapter, we present an overview of the IEEE 802.16 architecture, highlighting the salient features and briefly describing key MAC and PHY components. Detailed descriptions of individual MAC and PHY components of the IEEE 802.16 architecture are given in later chapters. The key aspects of the IEEE 802.16 architecture are summarized as follows:

- ***Flexible and extensible with common MAC:*** IEEE Std 802.16 defines a common MAC that works with multiple PHY technologies. The MAC is flexible enough to support many existing and future PHY technologies and other MAC extensions, as needed.
- ***Modular:*** Both the IEEE 802.16 MAC and PHY are very modular, having a set of mandatory and optional features that can be used to realize a variety of fixed and mobile configurations. Many of the optional features are negotiable between an SS and BS.
- ***Multiple network topologies:*** The support for single-hop and multihop network topologies such as PtP, PMP, and mesh have been built into the IEEE 802.16 MAC and PHY. This enables efficient implementation of any of the supported wireless topologies.
- ***Multiple antenna technologies:*** The IEEE 802.16 MAC and PHY work with standard antenna technologies such as omnidirectional or directional/sectorized antennas, in addition to multiple antenna technologies such as adaptive antennas and MIMO.
- ***MAC CSs:*** The IEEE 802.16 MAC is capable of transporting various encapsulated protocol payloads such as ATM, IP, and Ethernet. The core of the IEEE 802.16 MAC is independent of the type of payloads by defining multiple CSs and a standard interface between the core MAC and any CS. This allows the protocol-specific functions such as classification and

header compression to be performed outside the core MAC. Additional CSs may be defined in the future without having to change the core MAC.

- ***Flexible retransmission policies:*** IEEE Std 802.16 provides support for multiple optional retransmission policies, including an SR ARQ and a hybrid ARQ (HARQ).
- ***Privacy:*** The IEEE 802.16 MAC supports multiple encryption and authentication methods and is capable of supporting additional privacy mechanisms in the future. This is accomplished by a modularized security sublayer that is separated from the core MAC.
- ***Subscriber-level adaptive PHY:*** The IEEE 802.16 PHY allows a variety of parameters to be changed on a per-connection or per-subscriber basis that can be used to provide varying levels of service by adapting to changing conditions.
- ***Integrated QoS:*** The integrated QoS support at the IEEE 802.16 MAC defines multiple types of QoS that can be configured and controlled from higher layer. The IEEE 802.16 QoS model defines behaviors for various QoS types that require a QoS scheduler at the MAC. However, the scheduling algorithm is left unspecified to provide full flexibility for the implementor.
- ***TDD and FDD support:*** The IEEE 802.16 MAC and PHY support both TDD and FDD, with some restrictions as outlined in later sections.

The rest of the book provides an overview of these key features of IEEE Std 802.16 and explains the modules of IEEE Std 802.16 and the flexibility they provide to system designers.

Reference Model

The IEEE 802.16 standard describes both the MAC and PHY for fixed and mobile BWA systems. There are two major components of the wireless broadband system: the data/control plane and the management plane. The data plane defines how information is encapsulated or decapsulated in the MAC and modulated or demodulated by the PHY. A set of control functions is needed to support various configuration and coordination functions. While some standards refer to this as "management" (e.g., MAC and PHY

management in IEEE Std 802.11), the IEEE 802.16 standard refers to this as the *control plane* to differentiate from the external management system. Every broadband wireless system requires management, including management of the classification, security, QoS, connection setup, and other functions.

Figure 4–1 shows the IEEE 802.16 reference model. The IEEE 802.16 MAC consists of three major components called *sublayers*. The three sublayers are the service-specific convergence sublayer (CS), the MAC common part sublayer (CPS), and the security sublayer.

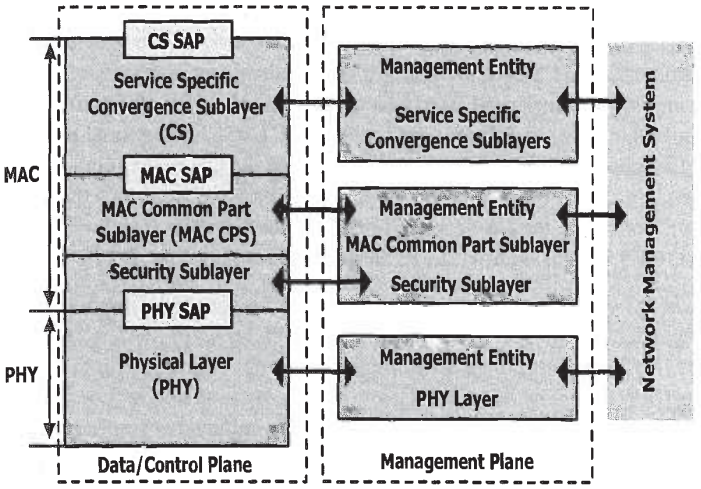


Figure 4–1: IEEE 802.16 reference model

As described in Chapter 3, SDUs need to be classified prior to QoS enforcement. Some higher layer protocols may have their own classification method and deliver preclassified SDUs to the MAC. For example, ATM has its own QoS framework, and QoS identifiers (IDs) are carried in each ATM cell. Similarly, IP also supports QoS protocols such as *differentiated services*. In a differentiated services network, IP packet headers carry differentiated services code points (DSCPs) that are used to identify the QoS class of each IP packet. Such preclassified traffic may be mapped directly into an

0056

IEEE 802.16 MAC connection or may be reclassified based on another set of rules. Other traffic may enter the IEEE 802.16 MAC unclassified, and full classification rules (e.g., based on MAC/IP address or port number) may be needed to classify these packets before applying QoS.

IEEE Std 802.16 currently supports a set of CSs to interface with IP, Ethernet, and ATM protocol layers. In the future, additional CSs may be defined. A service-specific CS performs all functions that are specific to the higher layer protocol it supports, such as classification. The Ethernet CS may classify SDUs based on the MAC address, virtual local area network (VLAN) headers, or IEEE 802.1Q user priorities. The IP CS may classify the SDUs based on the IP addresses, port numbers, protocol types, or other IP-specific parameters. The CS classification process is also responsible for associating a service flow identifier (SFID) and a connection identifier (CID) to the higher layer SDUs, prior to delivering it to the MAC CPS. The optional payload header suppression (PHS) is also performed at the CS, since it is very much dependent on the type of higher layer used. Chapter 5 describes the CS in detail. Following the CS-specific functions, the CS PDU is delivered to the MAC CPS, which is oblivious to the internal format of the CS PDU.

The MAC CPS is responsible for performing the core MAC functions that are independent of the specific CS. The MAC CPS receives MAC service data units (MSDUs) from the CS and transforms them into MAC protocol data units (MPDUs). The MAC CPS provides the medium access, connection management, and QoS functions. The security sublayer is responsible for providing encryption, decryption, authentication, and secure key exchange functions.

The MAC CPS interfaces with the PHYs through the PHY SAP. The MAC CPS may receive MSDUs from multiple MAC CSs. Although the MAC CPS defined in IEEE Std 802.16 can support many PHYs through a PHY-specific PHY SAP, one MAC CPS instance is expected to support only one specific PHY in a specific implementation, whereas it may support multiple MAC CSs in the same implementation. It should be noted that the PHY SAP is not explicitly defined for any of the PHYs in the standard. The QoS scheduling and allocation of bandwidth at the MAC CPS depends on the specific PHY type and parameters, although the MAC itself is PHY independent. The IEEE

802.16 MAC is flexible enough to support additional PHYs that may be defined in the future.

As shown in Figure 4–1, the management plane consists of four management entities corresponding to the CS, CPS, security sublayer, and the PHY. The IEEE 802.16 standard does not define the details of the management plane shown in Figure 4–1 as it is outside the scope of IEEE Std 802.16. However, specific interfaces and messaging may be standardized to support management functions and an IEEE 802.16 amendment. The IEEE 802.16g amendment currently under development is focused on defining management place procedures and services. The approved IEEE 802.16f amendment [B25] defines a MIB for the fixed wireless access systems. Some of the MAC and PHY control messages currently defined in IEEE Std 802.16 may also be used to manage IEEE 802.16 systems through an external management system.

BASE STATION (BS) AND SUBSCRIBER STATION (SS)

The IEEE 802.16 system architecture consists of two logical entities, the BS and SS. Both the BS and the SS have instances of the IEEE 802.16 MAC and PHY, in addition to other support functions. However, specific functions performed by the MAC or PHY differ depending whether it is a BS or an SS, and the IEEE 802.16 standard defines the BS- and SS-specific behavior in detail. In PtP and PMP networks, the BS and SS are in a *master-slave* relationship, where the SS must obey all medium access rules enforced by the BS. The SS in some configurations is referred to as the *customer premises equipment* (CPE) when the SS (or any part of it) is physically located within the customer's premises. Throughout this book, the BS and SS are considered logical entities.

In summary, the BS is responsible for the following functions:

- Enforcing basic MAC and PHY parameters such as frame size, ATDD, and configuration of system parameters
- Performing bandwidth allocation for DL (per connection) and UL traffic (per SS) and performing centralized QoS scheduling, based on the QoS/service parameters configured by the management system and the active bandwidth requests (BW requests) received from the SS

- Communicating the per-frame schedule to all SSs and supporting other data and management broadcast and multicast services
- Transmitting/receiving data and control information to/from one or more SSs within the same frame
- Performing connection admission control and other connection management functions
- Providing other SS support services such as ranging, clock synchronization, power control, and handoff

The SS is responsible for the following functions:

- Identifying the BS, acquiring PHY synchronization, obtaining MAC parameters, and joining the network if necessary
- Establishing basic connectivity, setting up additional data and management connections, and negotiating any optional parameters as needed
- Generating BW requests for connections that require such requests be generated, based on the connection profiles and traffic
- Receiving broadcast/multicast PDUs and unicast PDUs and forwarding them to the appropriate modules
- Making local scheduling decisions based on the current demand and history of BW requests/grants, when a BS allocates bandwidth for the SS
- Transmitting only when instructed by the BS to do so or the SS has some information that qualifies for transmission in one of the slots that may cause "contention" (e.g., ranging and BW requests in contention or broadcast allocations)
- Unless in sleep mode, receiving all schedule and channel information broadcast by the BS and obeying all medium access rules, transmitting data only when the BS allocates slots
- Performing initial ranging, maintenance ranging, power control, and other housekeeping functions

The mobile station (MS) defined in the IEEE 802.16 mobility extension (IEEE Std 802.16e) requires support for additional SS-specific functions such as mobility management, handoff, and power conservation.

One of the basic differences between the BS and SS in a PMP configuration is that the BS, which acts as a centralized controller and a centralized distribution/aggregation point, has to coordinate transmissions to/from multiple SSs, whereas the SS need to deal with only one BS. All traffic originating from an SS, including all SS-to-SS traffic, must go through the BS. Therefore, in a typical IEEE 802.16 system, the BS has to have additional processing and buffering capability (compared to a typical SS) to support a reasonable number of SSs.

The IEEE 802.16 mesh also includes the notion of BS and SS. However, the functionalities of these logical entities in mesh mode are slightly different from those of the BS and SS in PMP. The details of the BS and SS functions in a mesh are described in Chapter 9.

CONVERGENCE SUBLAYER (CS) ARCHITECTURE

The CSs enable the transport of data adhering to some other protocol specification, e.g., ATM, Ethernet, or IP, over an IEEE 802.16 link in a transparent fashion. The CS hides the details of the payload protocol from the IEEE 802.16 MAC, making it "protocol agnostic." Multiple CSs can coexist simultaneously, sharing the same MAC.

The CS for any protocol has to perform the following functions in the transmitter:

- Receive the payload protocol PDU from a higher protocol layer
- Map the payload protocol PDUs to the appropriate MAC service flow
- Optionally compress redundant payload protocol headers
- Deliver the processed packet to the MAC for transmission

In the receiver, the actions taken by the CS are as follows:

- Receive the MSDU
- Restore any compressed payload protocol headers
- Deliver the payload protocol PDU to the higher layer

The payload protocol PDU is mapped to the service flow by a set of configured rules called *classifiers*. The information a classifier considers

depends on the protocol being transported. In the case of Ethernet and IP, it is possible to aggregate frames or packets to the same service flow for transport over the air in the CS. The CSs are described in more detail in Chapter 5.

FRAMING AND DUPLEXING

As described in the previous sections, the CS converts upper layer PDUs into MPDUs, and the MAC CPS is responsible for transporting these MPDUs. The data encapsulated in IEEE 802.16 MPDUs are eventually carried in PHY frames. Multiple MPDUs destined for multiple connections and destinations may be carried within a single PHY frame. The duration of the PHY frame is fixed for a specific system configuration and does not change during the operation of the system, as shown in Figure 4–2. The IEEE 802.16 standard supports multiple PHY frame durations for each of the PHY types supported. The amount of higher layer information carried within a single PHY frame of specific duration depends on many factors, including the modulation and coding used, the size of the MPDUs, and other parameters that impact the per-frame overhead. As described in later chapters of this book, some of the per-frame overhead will be fixed for a specific PHY type, while some of the overhead will vary depending on parameters such as the number of SSs and number and types of PDUs carried within a frame.



Figure 4–2: IEEE 802.16 framing

IEEE Std 802.16 supports both TDD and FDD, with FDD currently allowed only in the licensed bands. By default, TDD systems can be either fixed TDD or ATDD. The standard supports full-duplex FDD and H-FDD operation, and both types of SSs can coexist in the same network. In a mixed network, the BS is responsible for making sure that a half-duplex SS does not receive overlapping allocations for UL and DL.

IEEE Std 802.16 is a framed air interface standard with fixed frames for both TDD and FDD. It is important to understand the following concepts before describing the TDD and FDD operation:

- ***Burst profiles:*** Each DL or UL burst is characterized by a set of parameters, called a *burst profile*, that includes information such as modulation type, FEC, preamble type, and guard times. The contents of the burst profile is PHY-specific.
- ***DL channel descriptor (DCD):*** DCD describes the DL PHY characteristics and is broadcast by the BS at periodic intervals. The DCD contains information such as the frame duration code and DL burst profiles. The DL burst profiles define DL interval usage codes (DIUCs) and associated PHY characteristics such as modulation and coding. The contents of the burst profile is PHY-specific. The DIUC is used to identify the burst profile of DL allocations in DL MAPs.
- ***UL channel descriptor (UCD):*** UCD describes the UL PHY characteristics. The UCD is also broadcast periodically by the BS. The UCD contains UL burst profiles that define UP interval usage codes (UIUCs) and associated PHY characteristics. In addition, the UCD contains backoff parameters to be used for the contention slots in the UL. All the UL and DL transmissions in IEEE Std 802.16 are scheduled; therefore, an SS cannot transmit data at its own discretion unless it is told to do so at a particular time offset. However, there are some cases where an SS may have to communicate some information to the BS without an explicit allocation. This includes a new SS that wants to register and join the network, an active SS that wants to maintain connectivity by requesting a robust DL burst profile, and an active SS that wants to send BW requests asynchronously. UL contention slots are defined to support these types of communication. The SSs are required to follow the specified backoff algorithms along with the parameters advertised by the BS in order to minimize the probability of collision in the contention slots.
- ***Downlink MAP (DL MAP):*** DL MAP describes the DL allocations. The allocations are specified in PHY-specific DL MAP elements. The DL MAP includes information on DL burst start time and DIUCs to identify the burst profile. While this is the only mechanism used in the WirelessMAN-SC and WirelessMAN-SCa PHYs to describe DL access,

WirelessMAN-OFDM and WirelessMAN-OFDMA PHYs support an additional format called downlink frame prefix (DLFP) to describe part of the DL allocations.

- ***Uplink MAP (UL MAP):*** UL MAP describes the UL TDMA allocations, which specify the exact time offsets, along with the burst profiles, with which the SSs will transmit data in the UL. The UL contention slots are also defined in the UL MAP. Similar to the DL MAP, the burst profiles are identified by UIUCs broadcast in UCDs.

As we will see in later chapters, there are many PHY-specific compressed or compact DL MAP and UL MAP formats defined in the standard to minimize the MAP overhead.

Physical slots (PS) and mini-slots

PSs and mini-slots are used in IEEE Std 802.16 to express quantities of time slot (bandwidth) allocation. PSs are the basic unit of allocation. The definition of PS depends on the specific PHY alternative. For the single-carrier PHYs, WirelessMAN-SC and WirelessMAN-SCa, PSs are defined to be four quadrature amplitude modulation (QAM) symbols in duration. For the OFDM-based PHYs, WirelessMAN-OFDM and WirelessMAN-OFDMA, PSs are defined to be $4/F_S$, where F_S is the sampling frequency.

Mini-slots are a single-carrier concept, allowing granularity coarser than the PSs in the UL when allocating bandwidth for WirelessMAN-SC and WirelessMAN-SCa. This coarser granularity is used to ensure that certain combinations of bandwidth and frame duration do not cause UL MAP entries to overflow. As such, the definition of a mini-slot varies from one system to another and is broadcast by the BS. A mini-slot is defined as 2^n PSs, where n ranges from 0 to 7. For the ODFM-based PHYs, UL opportunities are allocated in terms of OFDM symbols, and mini-slots are not used.

TDD framing

The structure of the IEEE 802.16 frame is illustrated in Figure 4–3. As shown in the figure, the fixed-length TDD frames are split into DL and UL portions, called DL and UL *subframes*. The TDD frame is repeated at regular intervals,

although the duration of the UL and DL subframes (i.e., TDD split) could be adjusted on a per-frame basis based on system parameters and run time parameters such as the amount of existing or expected DL/UL traffic and QoS. The UL and DL subframes carry one or more MPDUs. The DL MAPs and UL MAPs are broadcast at the start of each frame. DCDs and UCDs, if present, are also broadcast at the start of a frame.

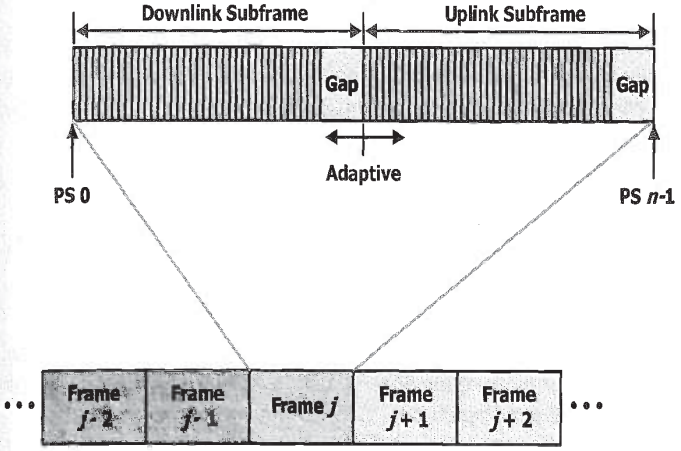


Figure 4–3: PMP TDD frame structure

TDD systems also require a receive-transmit transition gap (RTG) between the UL and subsequent DL subframe (of the next TDD frame). The duration of this gap varies, depending on the PHY type. This gap is used by the BS to transition from receive to transmit mode and by the SS to transition from transmit to receive mode. Similarly, a transmit-receive transition gap (TTG) is required between the DL and UL.

The DL subframe typically starts with information required for PHY synchronization, such as a preamble. The details of the preamble requirements for specific PHY modes are described in the PHY-specific

0060

chapters, Chapter 10, Chapter 11, and Chapter 12. The rest of the DL frame consists of multiple DL bursts.

A DL burst may consist of broadcast and/or unicast messages. Since the receiving SSs need to properly decode each burst, the DL MAP is broadcast at the beginning of the DL subframe. The UL MAP is also broadcast at the beginning of the DL subframe. As we will see later, the UL MAPs may refer to the UL portion of the current frame or a future frame such as the next frame, depending on a specific PHY implementation requirements. Other information related to the MAPs, such as the channel descriptors (DCD and UCD), may also be broadcast along with the DL MAPs and UL MAPs. The ATDD is controlled by a BS through these DL MAPs and UL MAPs, which determine the exact DL/UL split on a per-frame basis.

Apart from this frame descriptor information, other broadcast, multicast, or unicast information, such as the higher layer protocol PDUs or MAC management PDUs, is carried in the DL bursts. Each of these PDUs has a MAC header and optional payload that may be protected by a CRC. The construction of the MPDU is described in Chapter 6.

In summary, the DL subframe consists of any information to be transmitted from a BS to one or more of the SSs. The DL subframe is transmitted with TDM or TDMA, depending on the PHY and duplexing requirements.

The UL subframe is the TDMA portion, which may be used by one or more SSs to transmit information to the BS. The UL subframe may consist of contention slots and UL PDUs. The number of contention slots per UL subframe is determined by the BS on a per-frame basis. The UL PHY PDU is similar to the DL PHY PDU, where each UL PDU requires its own PHY synchronization information. However, unlike the DL PHY PDU, each UL PHY PDU can contain only one UL burst. The UL burst is similar to the DL burst, which is used to transport MPDUs and MAC management PDUs with a specific burst profile.

FDD framing

The structure of the FDD frame is shown in Figure 4–4. The structures of the DL and UL subframes are very similar to that of the TDD. FDD also uses

fixed frames. The primary difference with TDD is that the DL and UL subframes can overlap in time, as they are transmitted at different frequencies. However, the DL and UL subframes need not necessarily overlap in time. For example, the SS or BS may be incapable of full-duplex operation and, therefore, be unable to transmit and receive simultaneously on separate frequencies. If one of the devices is only half-duplex, the DL and UL subframes are not permitted to overlap in time, as shown in Figure 4–5. The UL MAP information must be provided to the SSs in time for them to initiate UL communication.

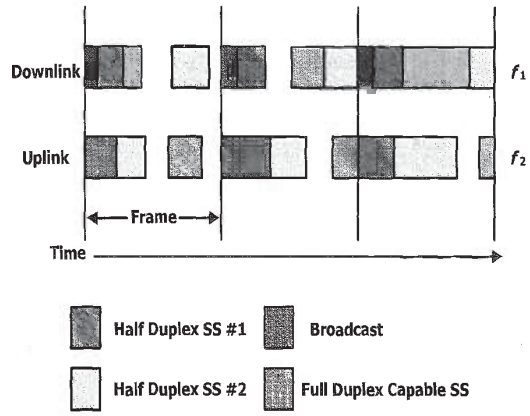


Figure 4–4: PMP FDD frame structure

Full-duplex FDD SSs do not require any gaps between DL and UL transmissions, as they are always listening and can transmit at any time allocated by the BS, as shown in Figure 4–6. However, H-FDD SSs, like TDD SSs, require a subscriber station receive-transmit transition gap (SSRTG) and subscriber station transmit-receive transition gap (SSTTG) between transmission and reception.

0061

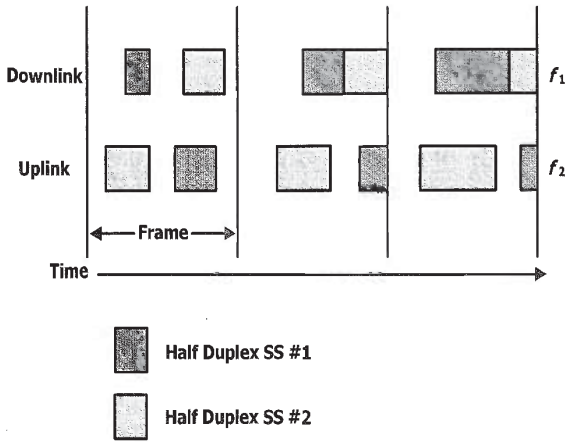


Figure 4–5: H-FDD framing

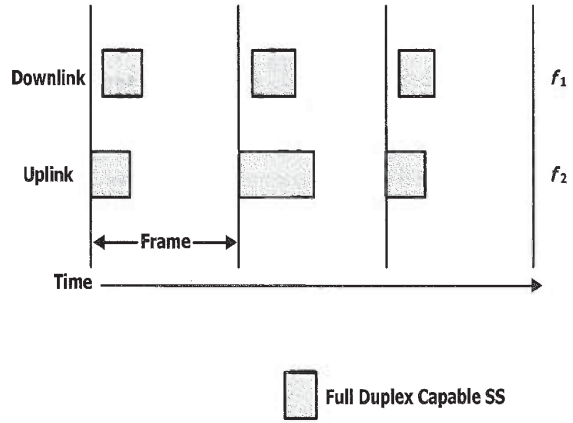


Figure 4–6: Full-duplex FDD framing

Time relevance

The DL MAPs and UL MAPs contain timing information for the SSs on when to receive and transmit. There are some rules on how far ahead in time DL and UL allocations can be broadcast by the BS. Obviously, real-time decisions can be made at the BS and SS if the DL MAPs and UL MAPs always refer to the current frame. However, implementations may not be able to construct and/or decode MAPs in real-time and make transmission and reception decisions based on the current frame's MAP. Therefore, IEEE Std 802.16 defines minimum and maximum time relevance rules that define the minimum and maximum time difference between the MAP broadcast and actual allocations.

The maximum time relevance for the UL MAP and DL MAP for the TDD framing is given in Figure 4–7. In TDD, the DL MAP always refers to the start of the current DL subframe. The UL MAP can be broadcast at most one frame ahead of time. For example, the UL MAP broadcast in Frame n cannot contain allocations for Frame $n + 2$. The minimum time relevance for the TDD framing is given in Figure 4–8. The UL MAP may refer to the start of the current UL subframe, if the SS implementation is capable of decoding the UL MAPs in the current frame and making transmission decisions.

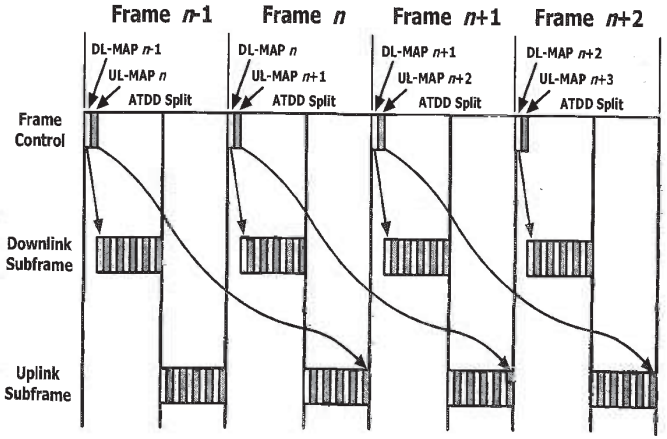


Figure 4–7: Maximum time relevance of UL and DL MAPs for TDD

62
0062

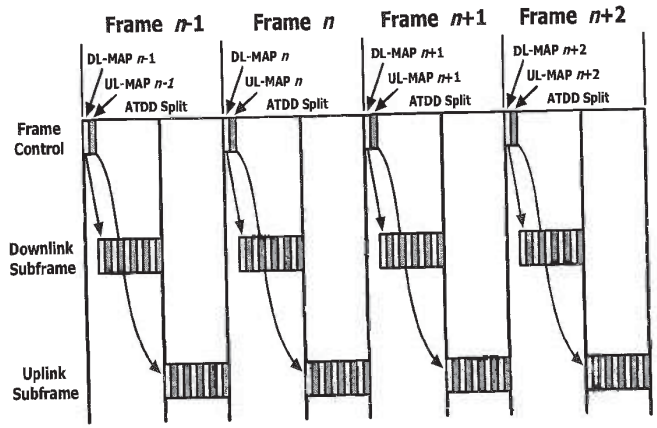


Figure 4–8: Minimum time relevance of UL and DL MAPs for TDD

The maximum time relevance of the FDD UL MAP and DL MAP is shown in Figure 4–9. Similar to TDD, the DL MAP always refers to the start of the current DL subframe, and the UL MAP can be broadcast at most one frame ahead of time. The minimum time relevance of the DL MAP and UL MAP for FDD is shown in Figure 4–10. The FDD UL MAP cannot refer to the start of the current frame, as there is not sufficient time for the SS to receive, decode, and transmit immediately. The minimum time relevance is controlled by two parameters. One is the round-trip delay that accounts for the propagation delay between the BS and the SSs. Another is the UL MAP processing delay, which defines the time required between the arrival of the last bit of the UL MAP in the SS and the time it takes to act upon the UL MAP. The UL MAP processing delay is PHY dependent.

More details on the time relevance and the rationale for the choices are explained later in Chapter 7.

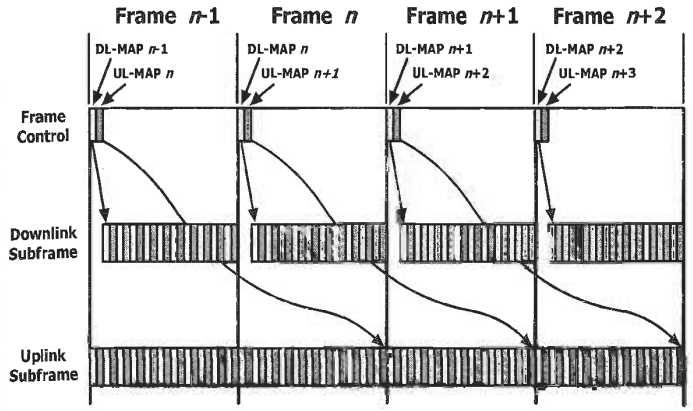


Figure 4–9: Maximum time relevance of DL and UL MAPs for FDD

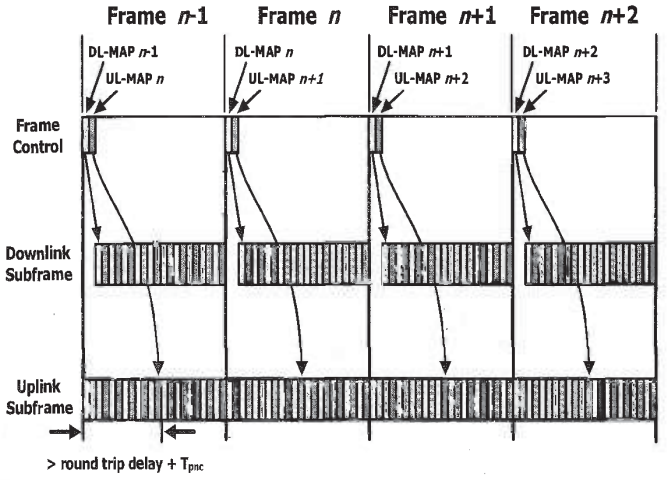


Figure 4–10: Minimum time relevance of DL and UL MAPs for FDD

0063

SUBSCRIBER-LEVEL ADAPTIVE PHY

The adaptive IEEE 802.16 PHY allows for each transmission burst, on each connection, to be described in terms of a burst profile. A subscriber may be characterized by one or more connections. A burst profile is identified by a system-wide code specified by the BS, called the *interval usage code* (IUC). The details of the supported parameters for the burst profiles of each PHY are described in the PHY-specific chapters, Chapter 10, Chapter 11, and Chapter 12. A burst profile is used for both UL and DL transmissions. An BS may use multiple burst profiles in the same frame to communicate with a single SS or multiple SSs.

The selection of a burst profile for a specific connection or subscriber should be determined by the BS based on a variety of constraints, including the QoS and channel conditions. The number of burst profiles supported in the UL or DL depends on the type of PHY. The BS assigns burst profiles based on various factors. For example, a bandwidth-efficient burst profile may be used under favorable link conditions, and a more robust, albeit less spectrally efficient, burst profile may be used when the link conditions are less favorable.

Fixed TDD vs ATDD

As discussed in Chapter 3, the choice between fixed TDD and ATDD depends on the system requirements and other design or implementation constraints. A fixed TDD configuration is typically easier to implement and may be sufficient for some systems with specific traffic patterns.

A minimum DL or UL portion is often required for basic TDD operation. For example, the DL MAPs or other broadcast information may be required in the DL. Depending on the system constraints, there may also be a minimum DL portion required in each TDD frame so that the SSs have sufficient time to switch from receive to transmit mode, after decoding the broadcast information, and transmit in the UL if necessary. Similarly, the UL portion may be required to include a minimum number of contention slots. An ATDD system may make decisions based on these constraints and other real-time constraints to determine the ATDD split. The minimum and maximum values for the ATDD split may also be specified as system parameters that can be

controlled from higher layers. ATDD allows a variety of traffic patterns to be supported in the system, including predominantly DL/UL, symmetrical, or bursty traffic.

FRAMED PHY

Early in the development of IEEE Std 802.16, a continuous PHY DL, as found in digital video broadcast (DVB) or data over cable service interface specification (DOCSIS®) systems, was suggested. A continuous PHY has the advantage that interleaving may be applied to increase system gain. In the end, however, the final specification did not include the continuous PHY, for a variety of reasons:

- TDD systems are naturally framed. For FDD and TDD system implementations to be as similar as possible, framed PHYs are more conducive.
- Because of the use of adaptive burst profiles to trade off robustness and efficiency in real time, there was no guarantee a burst would be long enough to take advantage of interleaving. Because of this, the system performance needed to be adequate without interleaving, negating most of the benefit of a continuous PHY.
- Adaptive burst profiles require an easy, natural resynchronization mechanism for SSs that could not stay synchronized during more complex PHY burst profiles. A framed PHY provides this natural resynchronization. A continuous PHY, on the other hand, assumes long-term synchronization and, therefore, complicates synchronization in the short term.
- IEEE Std 802.16 applies true QoS in both the UL and the DL. The framed PHY provides naturally occurring scheduling intervals for some QoS engines to operate at peak performance.

MAC EFFICIENCY

The efficiency of a MAC is determined by a variety of factors. In general, the overhead associated with medium access, retransmissions, and per-MPDU headers are key factors.

The medium access overhead (i.e., the time wasted in trying to access the medium) is almost zero in IEEE Std 802.16, as the centrally scheduled TDM/TDMA MAC can optimally schedule the transmissions without any unused bandwidth. The centrally scheduled MAC differs from distributed medium access methods such as carrier sense multiple access/collision avoidance (CSMA/CA), where nodes make independent transmit decisions, sensing the medium and backing off upon detection of another transmission or collisions. On the other hand, the centrally scheduled MAC introduces additional overhead since BW requests have to be communicated to the central scheduler.

Another source of inefficiency in a MAC is the overhead associated with the retransmission algorithm and associated acknowledgments. If every MPDU is acknowledged immediately, it provides instantaneous feedback and allows automatic retransmission decisions. However, the overhead associated with individual acknowledgments for every MPDU is very high. Moreover, it would not be possible to differentiate between the loss of the MPDU and the loss of an acknowledgment. One of the problems with immediate acknowledgment is that it cannot efficiently deal with PDU loss due to fading that lasts longer than the duration of a PDU. It may be desirable not to retransmit the frame immediately after the detection of the lost MPDU, as the probability of loss is still high for the retransmissions. IEEE Std 802.16 addresses these issues by supporting an SR retransmission algorithm, with efficient acknowledgment mechanisms such as selective and cumulative acknowledgments.

The per-PDU MAC overhead is minimized in IEEE Std 802.16 with the use of small headers. The basic MPDU header (i.e., generic MAC header) is only 6 bytes long and can be used to carry a MAC payload of up to 2041 bytes (when no CRC is included). The generic MAC header supports basic MAC transport functions, and any additional functions such as fragmentation and packing require additional information to be carried in MAC headers. IEEE Std 802.16 defines a set of subheaders, which are extensions to the generic MAC header and may optionally be included in an MPDU, to support such additional functions.

Some real-time applications, such as voice, generate smaller packets. The majority of the data networks today use TCP/IP, a connection-oriented protocol that requires small (typically 40 bytes and 52 bytes with the timestamp option) acknowledgment packets in the opposite direction of the data flow. A TCP acknowledgment is typically sent for every other TCP packet. These small packets are one of the sources of inefficiency in many MAC protocols, if the MAC allows only one SDU per MPDU. Since each MPDU requires a minimum MAC header and PHY header overhead, it is desirable from the efficiency perspective to have the ability to send more than one SDU in a single MPDU. In IEEE Std 802.16, this is achieved by "packing" multiple SDUs into a single MPDU for efficiency.

Another fundamental feature of IEEE Std 802.16 is dynamic fragmentation, where an SDU may be fragmented dynamically based on the real-time conditions, with no requirement for predetermined, fixed-size fragments. Dynamic fragmentation can be used to support efficient QoS as well as to efficiently use the available bandwidth. When the centralized scheduler allocates UL bandwidth for an SS, under no circumstances is the SS allowed to transmit beyond the allocated interval. Dynamic fragmentation allows the scheduler to fit the available frames into these allocated intervals and not leave slots unused.

The concept of ATDD in IEEE Std 802.16 may also be used to efficiently manage the available airtime by reducing wastage in TDD systems. Without ATDD, UL or DL subframes may be underutilized in some frames, as the instantaneous traffic may be asymmetric even in a system with many symmetric flows.

MESH

The IEEE 802.16 protocol includes an option that allows range extension by allowing SS-to-BS communication via one or more intermediate SSs. The IEEE 802.16 specification defines only the MAC-level protocol and assumes that routing is performed on the IP layer above the MAC. Also it is not possible to run the mesh protocol simultaneously with the main PMP protocol.

0065

The IEEE 802.16 mesh protocol offers two possibilities for scheduling transmissions. When centralized scheduling is used, the mesh BS collects data on transmission from the mesh SSs in its domain and computes the schedule for the domain. The schedule is then communicated to the mesh SSs using a special protocol. With distributed scheduling, the mesh SSs compute their own schedule based on information they collect from a two-hop neighborhood.

DIRECTED MESH

Directed mesh is a form of mesh where very directional antennas or beams are used to form a type of repeater network where a node (a combination of some SS and BS functions) acts as a repeater to one or more other nodes in the network. It may physically steer an antenna toward other nodes, or it may use beam forming to create a narrow beam directed toward other nodes. It may do so on the same frequency or on a different frequency. The narrow beams help provide significant frequency reuse and interference avoidance but at the expense of antenna system complexity and network entry complexity. The additional frequency reuse and interference tolerance can in many cases make up for the additional cost caused by this complexity. More details are provided in Chapter 8.

QUALITY OF SERVICE (QOS)

The QoS mechanism supported by IEEE Std 802.16 includes support for DL QoS and UL QoS. The IEEE 802.16 QoS model closely resembles the models defined in DOCSIS (see [B46]), with additional support for wireless-specific issues. As explained in Chapter 3, the QoS scheduling algorithm for wireless QoS has a different set of challenges. Like the DOCSIS and ATM QoS models, the IEEE 802.16 standard defines a QoS behavior and leaves the definitions of the actual QoS scheduler to the implementor.

Any IEEE 802.16 MAC supporting the defined QoS classes should have an integrated QoS scheduler. It is up to the vendor to incorporate a scheduling algorithm that meets the specified QoS requirements. The IEEE 802.16 QoS scheduler is different from a typical packet scheduler that works above the

MAC, as the IEEE 802.16 scheduler resides within the MAC, controlling various MAC parameters and reacting to changing channel conditions.

The IEEE 802.16 QoS support consists of the following:

- Four different QoS classes in the UL and DL
- Support for per-connection QoS and up to 64k connections in each direction
- Generic and PHY-specific mechanisms for UL BW requests and grants
- Flexible QoS models, where most of QoS scheduling can be done centrally at the BS

Since UL bandwidth is always granted per SS, the SS is responsible for making local scheduling decisions and making use of the granted bandwidth fairly and efficiently. In IEEE Std 802.16-2001, per-connection UL bandwidth grants were also optionally allowed. This allowed the BS to schedule UL transmissions on a per-connection basis and support per-connection burst profiles. However, due to the increased complexity at the BS scheduler and lack of flexibility at the BS, this option was later deleted and is no longer specified in IEEE Std 802.16-2004 [B20]. Per-SS grants may have an impact on the system performance when the BS and SS have schedulers with significantly different behaviors.

SECURITY SUBLAYER

The IEEE 802.16 security sublayer consists of two components: an encapsulation protocol for encrypting MAC payloads and a key management protocol called *privacy key management* (PKM). The encapsulation defines the encryption and data authentication method and the rules specifying how to apply these to the MPDU payload. The MAC headers are sent unencrypted. The subheaders, however, are encrypted as they are considered part of the payload. The majority of the MAC management messages are not encrypted, but integrity protection is afforded to management messages to prevent theft of service.

Encapsulation protocols include a data encryption standard (DES) in cipher block chaining (CBC) mode, which only encrypts the data but offers no

0066

integrity protection. The CBC initialization vector (CBC-IV) is derived from the frame number, and thus there is no data expansion as a result of applying this protocol. Due to use of the DES algorithm and the way it is used, this encapsulation does not offer a high level of security.

Much better security can be achieved by use of the advanced encryption standard with counter with CBC-MAC (AES-CCM) mode of encapsulating MPDU payloads. This method is essentially the same as the one used in IEEE Std 802.11i™. In addition to strong encryption, it also affords data integrity protection. The drawback is that the size of each MPDU is increased by 12 bytes.

The IEEE 802.16 key management protocol supports secure distribution of keys from BS to SS. The BS may use the key management protocol to enforce conditional access to the network services. The SS uses this PKM protocol to authorize traffic and obtain keys from the BS and performs periodic reauthorization and key refresh. Per-SS unique X.509 digital certificates (see [B19]) and RSA public-key encryption algorithms are used to perform key exchanges between SS and BS. The IEEE 802.16 MAC provides MAC management support for transporting PKM messages between BS and SS. The PKM protocol uses public-key cryptography to establish a shared secret between SS and BS, and the shared secret is used to secure subsequent PKM exchanges.

The IEEE 802.16e amendment introduces support for Extensible Authentication Protocol (EAP) and a new version of the key management protocol (PKM Version 2) that takes into consideration the issues brought into play in a mobile network. Some of the changes include explicit support for mutual and unilateral authentication, where the BS may authenticate the SS, but the SS may not have to authenticate the BS.

AUTOMATIC REPEAT REQUEST (ARQ)

The initial days of the IEEE 802.16 standard development saw support for three classes of ARQ, based on variations of stop-and-wait, GBN, and SR. There was consensus early in the process to not consider a MAC-level stop-and-wait ARQ algorithm. A lengthy debate ensued between the proponents of

GBN and SR variants. The primary argument in favor of GBN was its simplicity, as the SR algorithm would require additional signaling and buffering of out-of-order MPDUs. However, the working group eventually decided to use a SR variant, considering its efficient use of the spectrum. As we will see later, the stop-and-wait algorithm did return, in the form of HARQ.

The group also discussed, at length, the ARQ retransmission unit. Different proposals were floated, one based on the fragment as a retransmission unit and another based on a byte or block. If a fragment is used as a retransmission unit, its size must be decided ahead of time, and once transmitted it cannot be refragmented during retransmissions. On the other hand, a byte- or block-based retransmission would be able to repartition a retransmitted MPDU in any byte or block boundary. While the fragment-based retransmission is less complex, the byte- or block-based approach provided more flexibility to the scheduler in retransmission allocations. Initially, a compromise was reached to use a fragment-based retransmission unit. It was also decided to support transmission of an MSDU in multiple fragments in the first transmission (as a single MPDU), with the additional overhead of ARQ fragmentation headers. This was allowed to support some flexibility in retransmissions, where the original set of fragments could be retransmitted in any order in any number of fragments. However, this decision was also changed in a later revision in favor of only block-based retransmission.

As a result of these machinations, IEEE Std 802.16 currently supports an SR ARQ algorithm that is optionally enabled. The retransmission unit is a *block*, where a block is a contiguous block of at least 1 byte from the SDU, with a maximum size of 2040 bytes per block. The SDU is segmented into a number of blocks, and each transmission or retransmission may consist of any number of blocks. The concept of blocks is used to implement a flexible retransmission algorithm, where the number of blocks in an MPDU may change between retransmissions. The SDUs of an ARQ connection can be thought of as a sequence of blocks, where all blocks, with the possible exception of the last block of an SDU, are of the same size. An MPDU is constructed with an integral number of blocks, which may span multiple SDUs, with appropriate packing subheaders (PSHs).

0067

If the MPDU did not reach the receiver, the retransmission may involve the MPDU as originally constructed, or a new MPDU may be constructed as a subset of the original MPDU (i.e., with fewer blocks). Alternatively, a new MPDU may be constructed by combining the original MPDU (or part of the original MPDU) and other contiguous blocks. The implementation complexity of the block-based retransmission may be reduced by using MPDUs as retransmission units; i.e., keeping the MPDU intact between retransmissions. Other considerations, such as the interaction between encryption and fragmentation, may also force implementations to choose an MPDU-based retransmission.

A compact bitmap-based feedback is used by the ARQ algorithm, which supports multiple types of acknowledgments: cumulative, cumulative with selective, and selective. An implementation may use a mix and match of these acknowledgment types to efficiently use the protocol. It is also important to reduce the overhead associated with ARQ feedback. Therefore, the IEEE 802.16 standard allows the feedback to be sent (in the reverse direction of the information flow) either as stand-alone management messages or as payload piggybacked with other data. The specific algorithm to allocate bandwidth for the ARQ feedback, especially in the UL, is not specified in the standard but left to the developer. For example, the SS may periodically request additional bandwidth for ARQ feedback, or the BS may periodically allocate extra bandwidth for SSs with active ARQ connections. Similarly, bandwidth stealing may be used to send ARQ feedback.

Hybrid automatic repeat request (HARQ)

HARQ is implemented with the cooperation of both MAC and PHY, as opposed to the MAC-level ARQ that is PHY independent. In IEEE Std 802.16, the optional HARQ is supported only in the OFDMA PHY. The HARQ mechanism creates a PHY burst with one or more MPDUs and a per-burst CRC. Four subpackets, each identified by a subpacket identifier (SPID), are generated from the encoded HARQ packet. In the DL, the BS sends one of the subpackets in a burst. Since the subpackets are generated with redundancy, the SS can correctly decode the original encoded packet without having to receive all four subpackets.

As soon as the first subpacket is successfully received, the SS attempts to decode the original encoder packet. The SS sends an acknowledgment to the BS if the decoding of the original encoder packet was successful. If an acknowledgment is received, the BS stops sending additional subpackets of the same encoder packet. If the decoding is unsuccessful, the SS sends a negative acknowledgment. In response, the BS transmits the next subpacket from the four. This process continues until the SS successfully decodes the encoder packet or all attempts fail. Unlike the SR protocol of the MAC-level ARQ, the HARQ is a stop-and-wait protocol with immediate or synchronous acknowledgments. The HARQ mechanism provides a dedicated PHY channel for the SS to transmit positive or negative acknowledgments after a fixed delay. Although the acknowledgments are synchronous, the retransmission can be nondeterministic.

Similar to other stop-and-wait protocols, the IEEE 802.16 HARQ is simple to implement, with no additional buffering at the transmitter or the receiver. The disadvantages of stop-and-wait protocols, such as possible interaction with MAC QoS, also apply to this HARQ protocol.

PHYSICAL LAYER (PHY)

IEEE 802.16 has two different single-carrier PHY specifications:

- WirelessMAN-SC for use from 10 GHz to 66 GHz
- WirelessMAN-SCa for use below 11 GHz

WirelessMAN-SC was the first PHY approved for IEEE Std 802.16. At its high operating frequencies, propagation characteristics dictate that LOS operation is a practical necessity. This severely limits any concept of mobility. As a result, OFDM provides little benefit in this frequency range. A single-carrier PHY provides a simpler and more efficient alternative.

Below 10 GHz, it is a different matter. While the WirelessMAN-SCa PHY does provide a simpler transmit chain and more efficient alternative to WirelessMAN-OFDM and WirelessMAN-OFDMA, it may require a substantially more complicated equalizer in the receiver to combat the additional multipath at these frequencies. This can be either a time-domain equalizer with a substantial number of taps or a frequency-domain equalizer

0068

(FDE). In the latter case, one can abstractly draw a parallel between OFDM and single-carrier, as the Fourier transform in the OFDM transmitter is transferred across the link to the receiver in the form of FDE. At one time during the debate in the working group, a proposal suggested the use of OFDM in the DL and single-carrier in the UL. This would have confined most of the PHY complexity to the BS. Although this combination could have led to attractively simple SSs, the concept was rejected at the time due to the lack of consensus.

Multipath mitigation

In order to mitigate the degrading effect of multipath, the use of an equalizer is recommended with the WirelessMAN-SC and WirelessMAN-SCa PHYs. The duration of the relevant multipath components will typically be multiple symbols long. The cyclic prefix method is used in the WirelessMAN-OFDM and WirelessMAN-OFDMA PHYs. It is not common for the multipath to exceed the duration of the symbols in OFDM and OFDMA PHYs since the symbol duration in these two PHYs is at least hundreds of times longer than those in WirelessMAN-SC and WirelessMAN-SCa PHYs. An equalizer is hence typically not necessary for OFDM and OFDMA PHYs.

MANDATORY AND OPTIONAL COMPONENTS

Before considering the details of the MAC and PHY in later chapters, this chapter concludes with an overview of the major IEEE 802.16 optional and mandatory features. The standard itself is quite complex, with many options. However, specific streamlined system profiles have been defined for vendor interoperability. These are included in Clause 12 of the IEEE 802.16 standard.

Whether a feature is optional, mandatory, or forbidden depends on a variety of factors. Some features, such as support for quadrature phase-shift keying (QPSK) modulation, are mandatory for both BS and SS, irrespective of the PHY type. Many other features depend on the choice of PHY. For instance, MAC-level ARQ is not applicable for the WirelessMAN-SC PHY and is optional for all other cases. Once the PHY is chosen, features may be conditionally mandatory; in other words, implementation of an option makes another feature mandatory. Additionally, adherence to a system profile may

require the implementation of certain options, thereby making them mandatory.

Table 4–1 list major features of IEEE Std 802.16, along with the allowed range of values and status as mandatory or optional.

Table 4–1: Key IEEE 802.16 features and their status

Feature	Parameters and mandatory vs optional status
Channel bandwidths	1.25 MHz to 28 MHz, depending on PHY
FFT size	• 2048 for WirelessMAN-OFDMA (128, 512, 1024 added in IEEE Std 802.16e) • 256 for WirelessMAN-OFDM • N/A for WirelessMAN-SC and WirelessMAN-SCa PHYs
Duplexing	• Mandatory; at least one must be supported • Only one type of duplexing can be enabled at any time; i.e., TDD and FDD are mutually exclusive
Support for H-FDD SS	Mandatory, if FDD is supported
Frame duration	• PHY-dependent options • Must choose a single frame duration
Guard intervals (cyclic prefix)	• WirelessMAN-OFDM and WirelessMAN-OFDMA PHY-dependent options • Must choose a single guard interval
Two-branch transmit space-time coding (STC)	Optional
AAS	Optional
Subchannelization—DL	• Mandatory for WirelessMAN-OFDMA • Optional for WirelessMAN-OFDM • N/A for WirelessMAN-SC and WirelessMAN-SCa

69
0069

Table 4–1: Key IEEE 802.16 features and their status (Continued)

Feature	Parameters and mandatory vs optional status
Subchannelization—UL	• Mandatory for WirelessMAN-OFDMA • Optional for WirelessMAN-OFDM • N/A for WirelessMAN-SC and WirelessMAN-SCa
DFS	Mandatory for operation in certain license-exempt bands and regulatory domains
Modulation	• QPSK and 16-QAM mandatory • 64-QAM mandatory in some cases, optional in others • Binary phase shift keying (BPSK) and 256-QAM PHY-dependent
Turbo codes	Optional
CRC	• Optional for WirelessMAN-SC • Mandatory for all other PHYs • CRC may be enabled or disabled on a per-service-flow basis
Data encryption using DES	Mandatory
Data encryption using 128-bit AES-CCM	Optional
BW requests using BW request header	Mandatory
Focused contention BW requests	• N/A for WirelessMAN-SC, WirelessMAN-SCa, and WirelessMAN-OFDMA • Optional for WirelessMAN-OFDM
CDMA BW requests	• N/A for WirelessMAN-SC, WirelessMAN-SCa, and WirelessMAN-OFDM • Mandatory for WirelessMAN-OFDMA
Ability to initiate dynamic service flow creation	• Mandatory for BS • Optional for SS

Table 4–1: Key IEEE 802.16 features and their status (Continued)

Feature	Parameters and mandatory vs optional status
Piggyback BW requests/ Grant management subheader (GMSH)	Optional
Unsolicited grant service (UGS)	Optional, but system must support at least some service type
Real-time polling service (rtPS)	Optional, but system must support at least some service type
Nonreal-time polling service (nrtPS)	Optional, but system must support at least some service type
Best-effort (BE) service	Optional, but system must support at least some service type
Service-specific convergence sublayer (CS)	• Mandatory • Must support at least one variant
Security sublayer	Mandatory
PMP topology	Mandatory
Mesh	• Optional for WirelessMAN-OFDM PHY • N/A for other PHYs
HARQ	Optional
PHS	Optional
Fragmentation	Mandatory
Reassembly	Mandatory
ARQ	• Optional • N/A for WirelessMAN-SC PHY
Classification	Mandatory, within constraints of supported CSs and service types
Unpacking	Mandatory
Packing	Optional

70
0070

BIT ORDERING

Bit ordering is an often confusing aspect of a protocol specification or implementation. IEEE Std 802.16 follows a simple bit and byte ordering. The fields of MAC messages specified as binary numbers are transmitted as a sequence of bits, starting from the most significant bit (MSB). This rule applies to all fields of the MAC messages and fields of the type-length-values (TLVs), CRC, and header check sequence (HCS). The order of the fields is the same as the order in which they appear in the tables of this book and the standard. The bytes of the SDUs or SDU fragments are transmitted in the same order they were received from the higher layers. Some of the MAC management messages also define strings, which are transmitted in the same order as written. The bits of string symbols and SDU bytes are always transmitted starting with the MSB. The same rule applies to multiple-byte fields, where the MSB is the highest bit. For example, bit 15 and bit 31 are MSB for 16-bit and 32-bit fields, respectively.

Chapter 5 Convergence sublayers (CSs)

Support for multiple protocol transport

The CSs perform the task, at the transmitter side, of transforming a PDU from a higher layer protocol into an IEEE 802.16 MSDU and assigning the MSDU to a particular connection. The CS at the receiver side is responsible for the inverse operation. In the creation of the MSDU, some of the PDU headers may be suppressed. Two CSs are defined in the IEEE 802.16 standard, one for carrying ATM and one for carrying packet-based protocols.

ATM CS

An ATM connection is uniquely identified by its virtual path identifier (VPI) and virtual channel identifier (VCI). A connection may be switched based on virtual path (VP), in which case the cells are forwarded based on the VPI value in the cell header onto the outgoing VP, or virtual channel (VC), in which case cells are mapped to an outgoing connection based on the VPI/VCI pair.

The ATM CS behaves differently depending on whether it is configured into VP-switched or VC-switched modes. When operating in VP-switched mode, the CID of the MAC connection is determined solely based on the VPI field. The VPI field can optionally be suppressed over the air, saving 8 or 12 bits per cell, depending on the flavor of addressing used. The QoS parameters of the MAC connection should be set to match those of the ATM connection at the time the MAC connection is established. The mapping from MAC CID to VPI in the receiver is established at connection setup.

In VC-switched mode, the MAC CID is determined by the VPI/VCI pair, and both of these fields can be suppressed over the air. Again, the QoS and reverse mappings are determined at MAC connection setup time. Since there are fewer than 2^{16} MAC CIDs available, the full range of VPI/VCI values cannot be simultaneously supported. In practice, this limitation should not come into

0071

play. Figure 5–1 shows the generic form of an ATM CS PDU. When PHS is not enabled, the ATM CS PDU header is identical to the ATM cell header.



Figure 5–1: ATM CS PDU format

When PHS is enabled, the ATM CS PDU header is modified depending on whether the service is VP-switched or VC-switched. In the VP-switched case, the resultant ATM CS PDU header is as detailed in Figure 5–2. In the VC-switched case, the resultant ATM CS PDU header is as detailed in Figure 5–3.

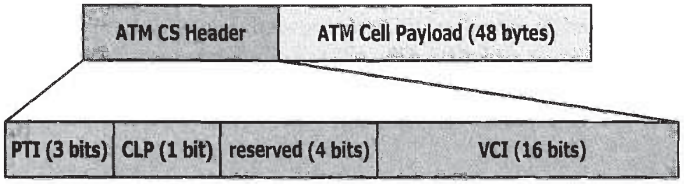


Figure 5–2: CS PDU format for VP-switched ATM connections

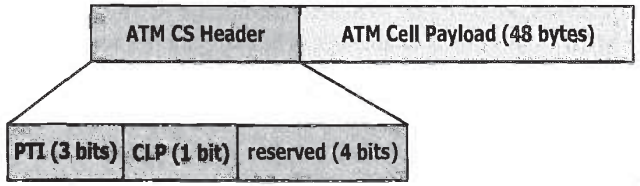


Figure 5–3: CS PDU format for VC-switched ATM connections

While the MAC CPS is allowed to pack ATM CS PDUs into MPDUs, it is not allowed to fragment ATM CS PDUs. This makes packing simpler than in the packet convergence sublayer (PCS) case because the packing process can take advantage of the fact that it has small, fixed-length packets.

However, this is complicated by PHS. The ATM CS can output packets of three different sizes. These correspond to full ATM cells with no PHS, VP-switched ATM cells with PHS, and VC-switched ATM cells with PHS. All ATM CS PDUs from the same service must use the same PHS mechanism. Only ATM CS PDUs from the same IEEE 802.16 MAC connection may be packed into the same MPDU.

PACKET CONVERGENCE SUBLAYER (PCS)

The PCS is designed to cope with any protocol utilizing packets for transporting data. Currently, the service flow signaling used for setting up the PCS supports only Ethernet and IP. There are, however, no fundamental limitations preventing additional support of other protocols such as multiprotocol label switching (MPLS) (see [B43]) either as a part of the standard or as a vendor extension. The PCS PDU is illustrated in Figure 5–4. Note that the PHSI field is always added when carrying Ethernet or IP, regardless of whether PHS is enabled for the connection. If PHS is used, the PHSI field indicates which PHS rule is to be applied by the receiver.

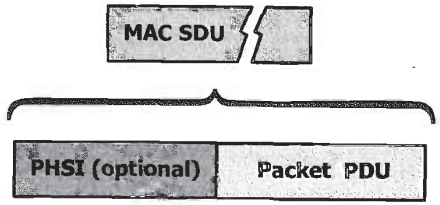


Figure 5–4: MSDU format for PCS

72
0072

Classification

The PCS contains a classification function that determines on which MAC connection a particular packet shall be carried and which PHS rule applies for that packet. The operation is illustrated in Figure 5–5. The exact rules of classification and the parameters by which the classifier is configured depends on the upper layer protocol and its design. Classifier parameters are configured during dynamic service signaling. Each classifier rule is also associated with a priority. Rules with a higher priority take precedence.

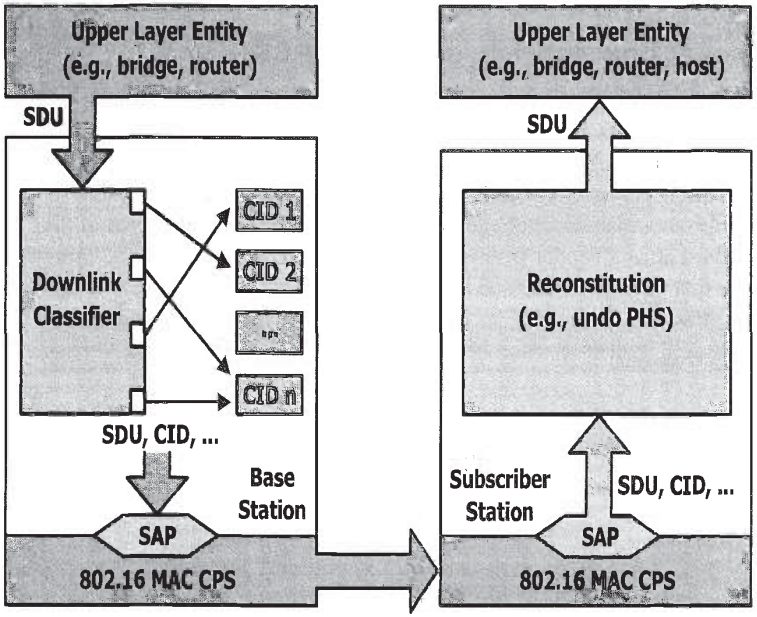


Figure 5–5: Classification and CID mapping (SS to BS)

Classification of Ethernet packets is based on the following fields in the headers, some of which may be absent:

- Ethernet destination MAC address
- Ethernet source MAC address
- Ethertype/IEEE Std 802.2™ SAP
- IEEE Std 802.1D™ User_Priority
- IEEE Std 802.1Q™ VLAN_ID

The operation performed by the classifier for all of the header fields is an exact match.

Classifying IP packets is more challenging. Often higher layer protocol headers need to be considered in order to enable a meaningful decision. The IP PCS classifiers can be configured to include TCP and User Datagram Protocol (UDP) headers in addition to the IP headers. Both IPv4 and IPv6 can be supported. The IP header fields that can be included in the classification are as follows:

- Source address
- Destination address
- Type of service (ToS)/differentiated services code point (DSCP)
- Protocol

The source address and destination address in an IP packet is matched to a set of prefixes stored in the classifier. Usually a longest prefix match is applied. As no particular classification algorithm is specified, the classifier rule priorities need to be set appropriately to ensure that the classifier behaves as expected. The ToS/DSCP value in the IP header is compared against a range of values after possibly masking away some of the bits. The Protocol field is specified by a set of numerical values checked for an exact match in the IP header.

In case the protocol carried over IP is one that uses "ports" such as TCP, UDP, and Stream Control Transmission Protocol, these can be included in the classification decision. The classifier can be configured to consider nonoverlapping ranges of port values in conjunction with a particular

Classification

The PCS contains a classification function that determines on which MAC connection a particular packet shall be carried and which PHS rule applies for that packet. The operation is illustrated in Figure 5–5. The exact rules of classification and the parameters by which the classifier is configured depends on the upper layer protocol and its design. Classifier parameters are configured during dynamic service signaling. Each classifier rule is also associated with a priority. Rules with a higher priority take precedence.

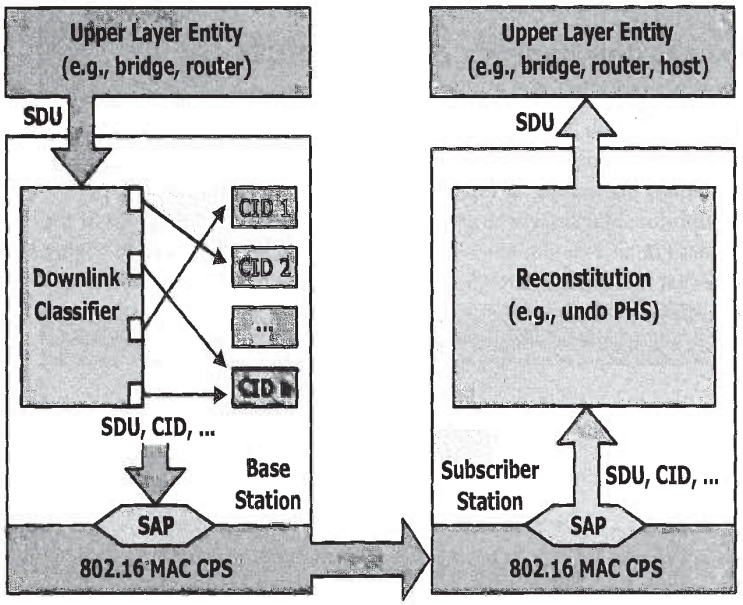


Figure 5–5: Classification and CID mapping (SS to BS)

Classification of Ethernet packets is based on the following fields in the headers, some of which may be absent:

- Ethernet destination MAC address
- Ethernet source MAC address
- Ethertype/IEEE Std 802.2™ SAP
- IEEE Std 802.1D™ User_Priority
- IEEE Std 802.1Q™ VLAN_ID

The operation performed by the classifier for all of the header fields is an exact match.

Classifying IP packets is more challenging. Often higher layer protocol headers need to be considered in order to enable a meaningful decision. The IP PCS classifiers can be configured to include TCP and User Datagram Protocol (UDP) headers in addition to the IP headers. Both IPv4 and IPv6 can be supported. The IP header fields that can be included in the classification are as follows:

- Source address
- Destination address
- Type of service (ToS)/differentiated services code point (DSCP)
- Protocol

The source address and destination address in an IP packet is matched to a set of prefixes stored in the classifier. Usually a longest prefix match is applied. As no particular classification algorithm is specified, the classifier rule priorities need to be set appropriately to ensure that the classifier behaves as expected. The ToS/DSCP value in the IP header is compared against a range of values after possibly masking away some of the bits. The Protocol field is specified by a set of numerical values checked for an exact match in the IP header.

In case the protocol carried over IP is one that uses "ports" such as TCP, UDP, and Stream Control Transmission Protocol, these can be included in the classification decision. The classifier can be configured to consider nonoverlapping ranges of port values in conjunction with a particular

protocol. A rule is not permitted to consider port values alone without considering the protocol.

Classification of IP packets is a relatively complex function. Implementations are not required to support classification on the full set of parameters defined by the IEEE 802.16 standard. When designing the protocol, however, it was agreed that the signaling should not be a limiting factor. Actually, the working group seriously considered the possibility of defining no classifier setup signaling in IEEE Std 802.16, instead relying on Internet Engineering Task Force (IETF) protocols. At the time, however, there was no single de-facto standard protocol for this purpose, and the protocols considered where deemed too "heavy."

Payload header suppression (PHS)

The PHS scheme defined for the PCS is inherited from DOCSIS (see [B46]). The scheme is protocol-independent and relatively simple, trading off some compression performance for ease of implementation. The PHS operation in the sender and receiver is illustrated in Figure 5–6. The BS and the SS negotiate the following PHS parameters constituting a PHS rule:

- ***Payload header suppression index (PHSI):*** The index to the PHS rule.
- ***Payload header suppression field (PHSF):*** A string of bytes containing the information allowing the receiver to reconstruct the suppressed information. The sender also keeps this information for purposes of verification (see PHSV below).
- ***Payload header suppression mask (PHSM):*** A string of bits, each corresponding to a byte in the PHSF. If a bit is set, the corresponding byte is suppressed in the sender, and the receiver will reconstruct the byte based on the content of the PHSF.
- ***Payload header suppression size (PHSS):*** A parameter indicating the total number of bytes to be processed by PHS. The value is always equal to the length (in bytes) of the PHSF.
- ***Payload header suppression valid (PHSV):*** A boolean value indicating whether the sender will compare the actual packet header with the version as it will be reconstructed by the receiver. If the two don't match exactly,

the headers of the higher layer packet are not suppressed, and the PHSI field in the PCS PDU is set to zero. Normally, this bit is set. However, some protocols and implementations may be able to cope with the loss of some of the header information, and consequently the verification can be disabled.

The PHS rule to be applied is determined by the sender based on the classification established in the PCS. The receiver, by inspecting the PHSI field, can determine the rule that was used to suppress a particular PCS PDU.

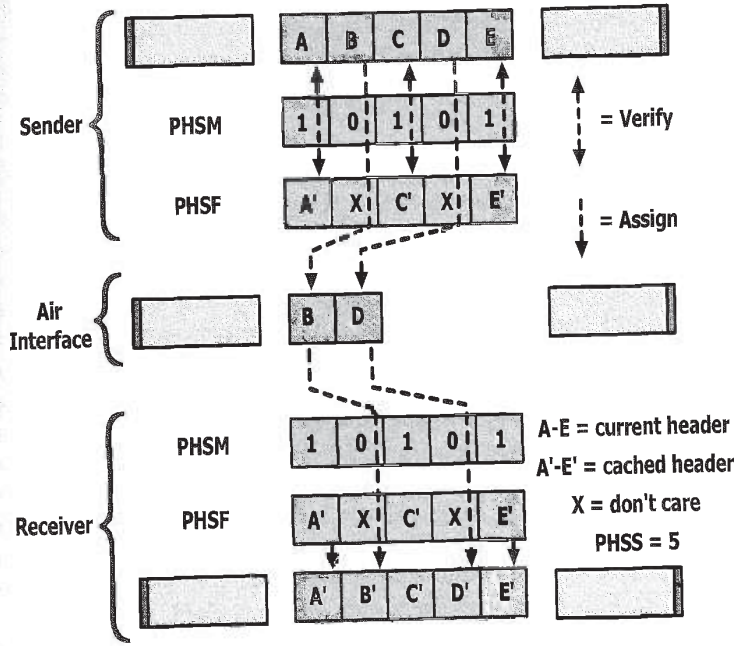


Figure 5–6: PHS operation, with masking

Chapter 6 MAC basics

Concepts, connections, formats, and headers

The IEEE 802.16 MAC specification covers much more than just the protocol for accessing the medium. As with cellular telephony standards, such as those within the Third Generation Partnership Project (3GPP), the IEEE 802.16 MAC specification also contains the associated protocols for radio resource control, radio link control, and security. The IEEE 802.16 MAC does not attempt to separate the control and user planes, but specifies both.

CONNECTIONS AND ADDRESSING

Each IEEE 802.16 MAC instance in an SS has a unique 48-bit MAC address as defined in IEEE Std 802®. An IEEE 802.16 SS may contain more than one instance of the MAC and can, therefore, have multiple 48-bit MAC addresses. One such example is an SS with multiple radios to support multihop or multichannel communication. Throughout this book, the term *MAC* refers to a single instance of the IEEE 802.16 MAC, unless otherwise specified; therefore, we assume there that each SS has a single 48-bit IEEE MAC address. However, this MAC address is used to uniquely identify an SS *only* during initial registration or authentication and as part of some management messages. For example, the initial ranging process, which must be completed before an SS can establish connectivity and gain access to a network, requires exchanging the SS MAC address for identification. On the other hand, it is important to understand that, unlike other wireless protocols such as IEEE 802.11, the 48-bit MAC addresses are *not* carried in every IEEE 802.16 MPDU to identify their source and destination. Instead, the connection-oriented IEEE 802.16 MAC makes primary use of the 16-bit CID to identify all information exchanged between BS and SS, including data, management, and broadcast data. The 16-bit CID is also the basic ID used in bandwidth allocations and bandwidth reservations. This not only reduces the overhead associated with carrying MAC addresses but also provides an easy way to

differentiate traffic. All MAC-level QoS functions, such as the classifier and QoS scheduler, use the CID to identify and differentiate traffic in order to maintain the service level and fairness between connections. Despite being connection- oriented, the protocol has been designed primarily to transport connectionless traffic, such as IP.

A maximum of 65 535 connections are supported. Some CIDs, or ranges of CIDs, are reserved for specific purposes. When a new SS wants to join a network, it starts without its own unique CIDs and cannot use any before the initial ranging is completed. A well-known default Initial Ranging CID is provided for this purpose. The details on when and how ranging messages are transmitted by an SS are described in Chapter 7 on MAC operation.

Since the 48-bit MAC address is not primarily used to identify an SS, a Basic CID is assigned to each SS after successful ranging. The Basic CID is used to transport basic MAC control/management messages, to identify an SS, and to communicate some per-SS information that is not specific to a connection. The later sections of this chapter and the next chapter describe various messages and procedures that make use of the Basic CID. Chapter 7 also describes the sequence and the minimum number of CIDs assigned to each SS. Additional CIDs for transporting data and management traffic may be negotiated during SS initialization or added dynamically during operation.

By definition, an IEEE 802.16 MAC connection is unidirectional, although in practice most of the services require a pair of connections, one in each direction. Some exceptions include one-to-many communication, such as multicast and broadcast, which are unidirectional. There are no limits to the number of connections between an SS MAC and BS MAC, subject to the maximum limits specified by the IEEE 802.16 MAC in general and subject to any practical limits that may be imposed by an implementation or to design limits based on scalability and manageability. The composition of the 16-bit CIDs used by the IEEE 802.16 mesh differs from the PMP CID concept as described above. Chapter 9 describes the mesh-specific CIDs and addressing concepts in detail.

The CID space is common between UL and DL. During the early stages of standards development, a separate CID space for UL and DL (i.e., 65 535 connections in each direction) was considered. However, the working group

decided against this in order to avoid any potential confusion and complexity. One exception to this rule is management connections, which share the same CID in both directions.

Service flows and service flow identifiers (SFIDs)

Service flows are used by the IEEE 802.16 MAC to efficiently support per-connection services such as QoS. When an SS is added to a system, SFIDs for this SS, along with associated QoS parameters, are provisioned, and each SFID is mapped onto a unique CID. Additional service flows may be defined, or the properties of an existing service flow may be modified depending on the activity and as the needs change. These changes are propagated to the SS or BS using the connection management messages. This model allows static configuration as well as dynamic modification of service flow properties. As we will see, not all CIDs have an associated service flow, as some of the CIDs are used for basic MAC management and control.

All higher layer protocols, connection-oriented or connectionless, are transported over an IEEE 802.16 connection. The actual association of the higher layer SDUs to IEEE 802.16 SFID/CIDs is performed by the CS, as described in Chapter 5. The IEEE 802.16 MAC connections are in no way related to any other notion of connection or lack thereof at the higher layers. It is possible to define a one-to-one mapping between a higher layer connection and an IEEE 802.16 MAC connection. It is also possible to map multiple higher layer connections into a single IEEE 802.16 MAC connection. Similarly, a mix of higher layer connections and connectionless traffic may be mapped onto a single IEEE 802.16 MAC connection. Such mappings are defined by the classifiers used in the IEEE 802.16 MAC CS described in Chapter 5.

CID allocation

The following types of CIDs are defined in IEEE Std 802.16:

- ***Initial Ranging CID:*** Initial ranging needs to be completed before an SS can establish connectivity to the BS and obtain provisioning parameters, such as other CIDs. Since an SS has no unique CID available before initial

ranging, it uses the known Initial Ranging CID for this purpose. The Initial Ranging CID should not be used by an SS if it has already obtained its Basic CID. Since both the BS and SS have to exchange ranging information as part of the initial ranging process, the Initial Ranging CID is reserved in both UL and DL.

- ***Basic CID:*** Each SS is assigned a Basic CID, during the ranging process, to exchange delay-intolerant and time-critical MAC management messages with the BS. The Basic CID may also be used to identify the SS for the purpose of managing or implementing per-SS functions. The same Basic CID is assigned to both the UL and DL connections used to transport basic MAC information to and from the SS. The use of the Basic CID in specific MAC management messages is described throughout this chapter and the next chapter.
- ***Primary Management CID:*** Each SS is assigned a Primary Management CID to transport delay-tolerant MAC management messages between SS and BS. The same Primary Management CID is assigned, during the ranging process, to both UL and DL connections for each SS. Appendix A shows the MAC management messages and associated connections used to transport these messages.
- ***Secondary Management CID:*** The secondary management connection is used to transport higher layer management messages, such as those supporting Simple Network Management Protocol (SNMP), Trivial File Transfer Protocol (TFTP), and Dynamic Host Configuration Protocol (DHCP). The Secondary Management CID is assigned to the SS if, and only if, it identifies itself as a "managed SS" during the registration process following initial ranging. The same Secondary Management CID is assigned in both UL and DL.
- ***Transport CID:*** Each SS is assigned zero or more transport connections after the completion of ranging and authentication. At least one Transport CID is necessary to transport any user information to or from an SS. The UL and DL Transport CIDs may be assigned independently, since the connections are unidirectional by definition. However, most applications using IEEE 802.16 unicast connections (e.g., TCP-based applications that require TCP acknowledgments in the reverse direction and VoIP that requires the transport of signalling and voice packets in both directions)

require bidirectional information exchange. Separate multicast Transport CIDs are not defined. A multicast connection is established by setting up individual connections with the same CID to each SS belonging to the multicast group. An SS may not even be aware that a specific Transport CID is multicast.

- ***AAS Initial Ranging CID:*** The AAS Initial Ranging CID is used by BSs that support AAS to allocate an initial ranging period for AAS SSs. It is also used by the AAS SSs to initiate the initial ranging process. A separate Initial Ranging CID is required for AAS SSs to prevent the non-AAS SSs from using the slots allocated for AAS SS ranging. Otherwise, the AAS Initial Ranging CID is used in a manner very similar to that for the Initial Ranging CID. Chapter 13 describes the AAS initial ranging process.
- ***Multicast Polling CID:*** The Multicast Polling CIDs are used to manage the bandwidth (i.e., transmission opportunity) request process. It is important to allocate some UL bandwidth in which the SS can request bandwidth for its connections. This is especially important if the SS has no active UL traffic, which, as we will see later, could allow piggybacked BW requests. It is neither scalable nor efficient for the BS to allocate individual transmit opportunities for all SSs in anticipation of BW requests from specific SSs. To more effectively address this problem, multicast polling groups are defined. An SS may be a member of one or more multicast polling groups. The BS can poll the multicast polling groups by allocating transmission opportunities for the multicast CID so that any SS that is a member of this multicast polling group may use this opportunity to request bandwidth. Since more than one SS may request bandwidth using a single transmit opportunity, collision of BW requests is possible. The multicast Transport CID is used to transmit information in the DL, and Multicast Polling CIDs are used to define UL BW request contention regions. No service flows can be associated with a Multicast Polling CID.
- ***Padding CID:*** The Padding CID is used to transport padding information by the SS or BS in UL and DL, respectively. Padding information is sent by the SS or BS when it has no other valid information to transmit in an allocated opportunity. The padding byte 0xFF is used in the payload of padding information.

- ***Broadcast CID:*** The Broadcast CID is used by the BS to broadcast MAC management information to all SSs in the DL. The Broadcast CID is not used in the UL. Apart from the well-known default Broadcast CID, additional Broadcast CIDs may be defined using one of the multicast CIDs to support a variety of additional broadcast functions.
- ***Fragmentable Broadcast CID:*** The Fragmentable Broadcast CID is used by the BS to transmit fragmented management broadcast information. Management messages sent by a BS in the regular Broadcast CID are not fragmentable. The Fragmentable Broadcast CID can be used by a BS to transmit management information to overcome this restriction. Management broadcast information is typically sent at the most robust modulation and coding so that all SSs can receive it. Therefore, fragmenting some long MAC management messages, in order to maintain fairness and QoS, is advantageous.

Table 6–1 shows the CID allocation in the IEEE 802.16 MAC and the well known CIDs, of which there are five: Initial Ranging CID, AAS Initial Ranging CID (used only by AAS-capable SSs), Padding CID, Broadcast CID, and Fragmentable Broadcast CID. A total of 253 CIDs are allocated to multicast polling groups. The rest of the 65 182 CIDs are reserved for use as a Basic, Primary Management, Secondary Management, and Transport CID, as defined in Table 6–1. The transport connections may include DL multicast connections. The value of m in Table 6–1 depends on the configuration of a specific deployment. Since the Basic and Primary Management CID ranges are defined in terms of m, and there is exactly one Basic and one Primary Management CID per SS, the choice of m is a system parameter that essentially defines the maximum possible number of SSs that can be supported.

Table 6–1: CID allocation and well-known CIDs

CID/Range	Purpose/Type
0x0000	Initial Ranging CID
0x0001 through m	Basic CID
(m + 1) to 2m	Primary Management CIDs
(2m + 1) through 0xFEFE	Transport and Secondary Management CIDs
0xFEFF	AAS Initial Ranging CID
0xFF00 through 0xFFFC	Multicast Polling CIDs
0xFFFD	Fragmentable Broadcast CID
0xFFFE	Padding CID
0xFFFF	Broadcast CID

MAC HEADERS AND SUBHEADERS

Recall from Chapter 3 that the PDU is the data unit exchanged between the peer entities of the same protocol layer. The IEEE 802.16 MPDUs are the units of data delivered by the MAC to the PHY for transmission and the unit of data delivered by the PHY to the MAC in the receiver. The IEEE 802.16 MAC receives MSDUs (e.g., encapsulated IP packets, ATM cells, Ethernet frames) from the CS above.

There are two classes of IEEE 802.16 MAC information transported between peer entities: stand-alone MAC headers and MPDUs. The 6-byte stand-alone header is the smallest possible information unit that can be transported between two nodes, with the exception of HARQ MAPs, which are sent without the generic MAC headers as we will describe later in this chapter.

Stand-alone MAC headers

Stand-alone MAC headers are used to transport compact control or signalling information between an SS and a BS. None of the stand-alone headers can be used to encapsulate any payload, as they are self-contained and used for a

0079

specific purpose. Therefore, it is, technically, a misnomer to call them "headers." All of the currently defined stand-alone MAC headers are 6 bytes in length, as is the generic MAC header used to carry payloads. Future IEEE 802.16 extensions may define additional MAC headers to support other functions.

All currently defined stand-alone MAC headers are relevant only in the UL. Therefore, only the BS is required to process these headers; any SS receiving a stand-alone header should discard it. A header type (HT) bit is used to indicate whether the header is a generic MAC header (HT = 0) or one of the stand-alone headers (HT = 1). Additional fields, such as the Type field, are used to further distinguish the type of generic or stand-alone header being transported or other information that may be present in the payload. In this book, reserved fields are indicated by RSV or Rsv.

BW request header

The 6-byte BW request header (Figure 6–1) carries a full BW request from an SS. Table 6–2 describes the interpretation of the BW request header fields. As shown in the table, there are two types of BW request header: *incremental* and *aggregate*. These are identified by the Type field. The BW request header is protected by a 8-bit HCS.

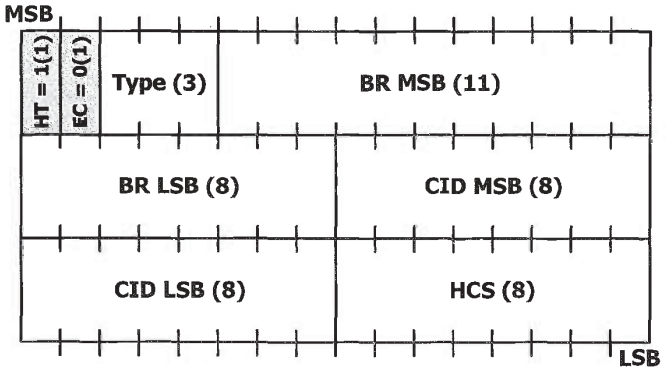


Figure 6–1: BW request header

Table 6–2: BW request header fields

Field	Length (bit)	Description
HT	1	*Header type:* Always set to 1, indicating this is not a generic MAC header.
EC	1	*Encryption control:* Set to 0, indicating no encryption.
Type	3	*Type:* Indicates the type of the BW request. Two types of BW requests are defined, although the Type field is used to demultiplex the different HTs. 000 - Incremental BW request 001 - Aggregate BW request
BR	19	*BW request:* Indicates the amount of UL bandwidth, in bytes, being requested for this CID. This does not include any PHY overhead.
CID	16	*Connection identifier:* The CID for which this BW request is being generated.
HCS	8	*Header check sequence:* The HCS protects the first 5 bytes of the MAC header. It is calculated as the remainder of the division (modulo 2) by the generator polynomial $g(D)=D^8+D^2+D+1$ of the polynomial D^8 multiplied by the first 5 bytes of the MAC header.

MPDU header

Each MPDU begins with a 6-byte generic MAC header, followed by the optional variable-size payload. The payload may consist of MAC subheaders, management messages, special payloads, MSDUs received from a CS, or padding information.

Generic MAC header

Figure 6–2 shows the generic MAC header used to encapsulate all IEEE 802.16 data and management PDUs. Table 6–3 describes the fields of the generic MAC header in detail.

0080

specific purpose. Therefore, it is, technically, a misnomer to call them "headers." All of the currently defined stand-alone MAC headers are 6 bytes in length, as is the generic MAC header used to carry payloads. Future IEEE 802.16 extensions may define additional MAC headers to support other functions.

All currently defined stand-alone MAC headers are relevant only in the UL. Therefore, only the BS is required to process these headers; any SS receiving a stand-alone header should discard it. A header type (HT) bit is used to indicate whether the header is a generic MAC header (HT = 0) or one of the stand-alone headers (HT = 1). Additional fields, such as the Type field, are used to further distinguish the type of generic or stand-alone header being transported or other information that may be present in the payload. In this book, reserved fields are indicated by RSV or Rsv.

BW request header

The 6-byte BW request header (Figure 6–1) carries a full BW request from an SS. Table 6–2 describes the interpretation of the BW request header fields. As shown in the table, there are two types of BW request header: *incremental* and *aggregate*. These are identified by the Type field. The BW request header is protected by a 8-bit HCS.

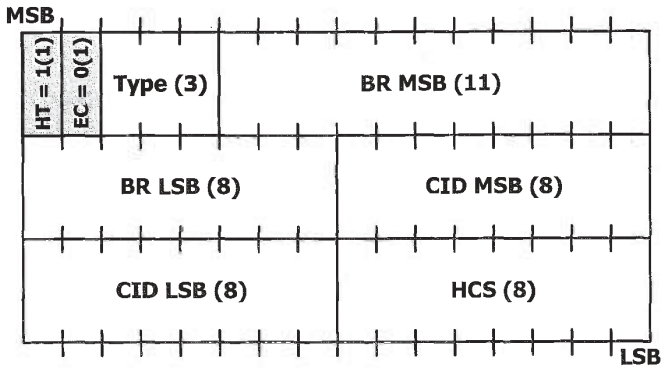


Figure 6–1: BW request header

Table 6–2: BW request header fields

Field	Length (bit)	Description
HT	1	*Header type:* Always set to 1, indicating this is not a generic MAC header.
EC	1	*Encryption control:* Set to 0, indicating no encryption.
Type	3	*Type:* Indicates the type of the BW request. Two types of BW requests are defined, although the Type field is used to demultiplex the different HTs. 000 - Incremental BW request 001 - Aggregate BW request
BR	19	*BW request:* Indicates the amount of UL bandwidth, in bytes, being requested for this CID. This does not include any PHY overhead.
CID	16	*Connection identifier:* The CID for which this BW request is being generated.
HCS	8	*Header check sequence:* The HCS protects the first 5 bytes of the MAC header. It is calculated as the remainder of the division (modulo 2) by the generator polynomial $g(D)=D^8+D^2+D+1$ of the polynomial D^8 multiplied by the first 5 bytes of the MAC header.

MPDU header

Each MPDU begins with a 6-byte generic MAC header, followed by the optional variable-size payload. The payload may consist of MAC subheaders, management messages, special payloads, MSDUs received from a CS, or padding information.

Generic MAC header

Figure 6–2 shows the generic MAC header used to encapsulate all IEEE 802.16 data and management PDUs. Table 6–3 describes the fields of the generic MAC header in detail.

0081

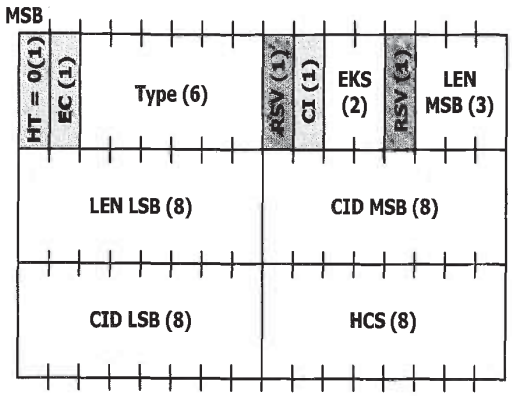


Figure 6–2: Generic MAC header

The type of payload carried within an MPDU is indicated by the CID and the Type field in the generic MAC header. If the CID is a management CID, the payload may consist of one or more MAC management messages. There are some restrictions on fragmentation and packing of some MAC management messages carried on certain management CIDs. If the CID is a Transport CID, the payload may consist of one or more MSDUs or MSDU fragments. The Type field is used to indicate the presence of various MAC subheaders and special payloads.

This generic MAC header may not be very friendly to software implementations, given the rather odd alignment of some fields, but it has all the fields necessary for basic MAC communication. Optional subheaders may be used to support additional MAC functions. Like the BW request header, the generic MAC header is also protected by an 8-bit HCS. If subheaders are included, they are considered part of the payload and not protected by this HCS. The payload may be protected by the optional CRC.

Table 6–3: Generic MAC header fields

Field	Length (bit)	Description
HT	1	*Header type:* Indicates if the header is a generic MAC header or a stand-alone header. Always set to 0 for the generic MAC header.
EC	1	*Encryption control:* Indicates whether the payload is encrypted. 0 = Not encrypted 1 = Encrypted
Type	6	*Payload type:* Indicates the type of subheaders or the type of special payloads present in the payload, if any. See Table 6–4 for the type encodings.
CI	1	*CRC indicator:* Indicates whether a CRC-32 has been included after the encryption, if applicable. 0 = No CRC included 1 = CRC included
EKS	2	*Encryption key sequence:* Indicates the index of the traffic encryption key (TEK) and initialization vector (IV) used to encrypt the payload. These bits are ignored if EC = 0.
LEN	11	*Length:* Indicates the length of the payload in bytes, including the MAC header and CRC if present.
CID	16	*Connection identifier:* Indicates the CID of the connection to which the payload belongs. See the description on packing for more details on packing payloads from multiple connections in the same MPDU.
HCS	8	*Header check sequence:* Same as described in Table 6–2.

Table 6–4 shows the interpretation of the 6-bit Generic MAC HT field. Each bit of the Type field corresponds to an optional subheader or a special payload type. If the subheader or the special payload is present in this MPDU, the corresponding bit position in the Type field is set to one; otherwise, it is set to zero. The description and usage of the subheaders and payload types are given in later sections.

Table 6–4: Generic HT encodings

Bit	Corresponding subheader or payload
0	*DOWNLINK:* Fast-feedback allocation subheader (FFSH). *UPLINK:* Grant management subheader (GMSH).
1	Packing subheader (PSH).
2	Fragmentation subheader (FSH).
3	Extended version of fragmentation and/or PSH, used when ARQ is used. The presence of the fragmentation or PSH is indicated separately. This bit is set to 0 if ARQ is not used.
4	ARQ feedback payload.
5	Mesh subheader (MSH).

MAC header demultiplexing

Table 6–5 shows how various MAC headers are demultiplexed by a BS based on the HT, EC, and Type header fields. The HT field is used to distinguish a generic MAC header from a stand-alone header. Some of the reserved Type field values and the EC field may be used in the future to define additional stand-alone MAC headers. Since 0xFF is the stuff byte used to fill allocated but unused slots, the MAC header formats are defined so that no MAC header starts with 0xF.

Table 6–5: MAC header demultiplexing

HT	EC	Type	Header
0			Generic MAC (see Table 6–4 for type encodings)
1	0	000	Incremental BW request
		001	Aggregate BW request
		010	Reserved
		011	Reserved
		100	Reserved
		101	Reserved
		110	Reserved
	1		Reserved

MAC subheaders

The MAC subheaders are used to extend the basic functionality supported by the generic MAC header. All subheaders are defined per PDU, except for the PSH, which is defined per SDU. As we will see in later subsections, the PSH is used to pack more than one SDU or SDU fragment into a single MPDU. Since the PSH serves as a superset of the FSH, the PSH and FSH are mutually exclusive. The FSH can be used only when the MPDU consists of only one SDU or SDU fragment.

As shown in Table 6–4, the Type field of the generic MAC header indicates the type(s) of subheaders present in the MPDU. Since the Type field indicates only the presence or absence of subheaders, an ordering of the placement of subheaders after the generic MAC header is enforced in order to properly decode the subheaders whenever multiple subheaders are present in a single MPDU.

The following subsections describe in detail the subheader formats and the MAC operations supported by these subheaders.

Fragmentation subheader (FSH)

FSHs are used to transport encapsulated MSDU or SDU fragments in a connection. Therefore, they carry sufficient information for the receiver to properly identify and reassemble fragments. Once a connection is provisioned with fragmentation enabled, all MPDUs transported on that connection should have a FSH, irrespective of whether a specific MPDU carries a fragmented or full SDU.

Three different types of FSHs are defined in IEEE Std 802.16, as illustrated in Figure 6–3 and Table 6–6.

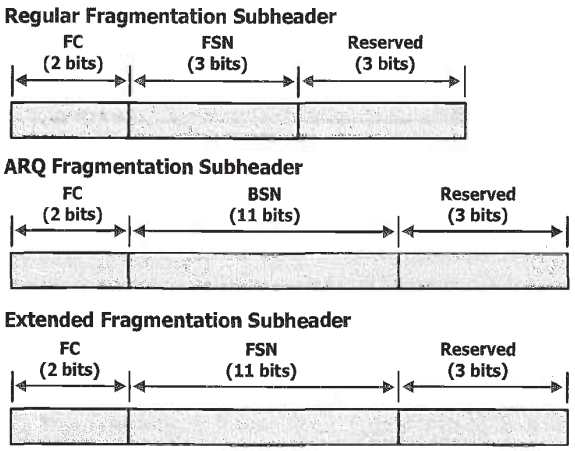


Figure 6–3: FSHs

Table 6–6: FSH fields

Field	Size (bit)	Subheader type	Comments
FC	2	Present in all FSHs	Indicates the type of fragment: 00 = No fragmentation 01 = Last fragment 10 = First fragment 11 = Continuing or middle fragment
FSN	3	Regular	Sequence number of this fragment. Increments by one, modulo 8, for every SDU or SDU fragment.
	11	Extended	Increments by one, modulo 2048, for every SDU or SDU fragment.
BSN	11	ARQ	Sequence number of the first block in this SDU or SDU fragment.

All FSHs include the 2-bit FC (fragment control) field that indicates the fragmentation state of the current fragment. The regular FSH is 1 byte in length, with a 3-bit fragment sequence number (FSN). The extended FSH is similar to the regular FSH, except that it is 2 bytes long and supports an 11-bit FSN space. The 2-byte ARQ FSH, on the other hand, uses a 11-bit block sequence number (BSN) instead of the FSN. The concept of blocks and the BSN are described elsewhere in this book.

Packing subheader (PSH)

Packing allows efficient transmission of multiple MSDUs or MSDU fragments in a single MPDU. As we will see later, the ARQ mechanism requires the SDUs or SDU fragments be transmitted out of sequence when retransmitting blocks lost during previous transmissions. The packing mechanism is designed to support such transmission of noncontiguous ARQ blocks or fragments. Three PSHs, similar to the FSHs, are defined, as shown in Figure 6–4.

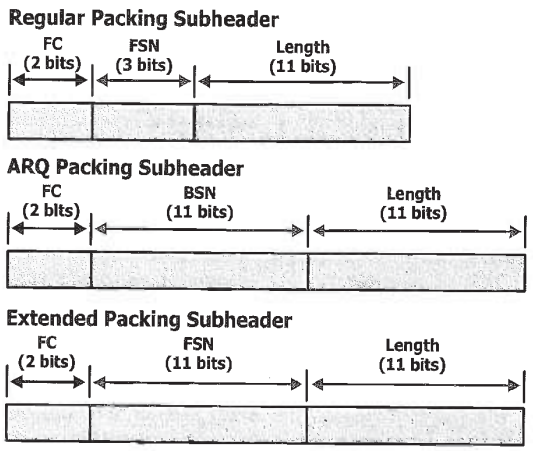


Figure 6–4: PSHs

The three PSHs parallel the corresponding FSHs, with an 11-bit Length field in place of the three reserved bits. This was the motivation for defining an extended FSH, instead of using the three reserved bits of the regular FSH to extend the FSN.

The Length field indicates the length of the unit of data being packed. This includes the length, in bytes, of the SDU/SDU fragment plus the PSH itself. The definitions of the rest of the PSH fields are the same as the corresponding fields in Table 6–6.

Grant management subheader (GMSH)

The 2-byte GMSH is used by the SS to request additional bandwidth and transmit other signalling related to bandwidth allocation. This BW request is essentially piggybacked onto an ongoing connection, in contrast to the stand-alone BW request using a BW request header. The IEEE 802.16 QoS model defines a set of rules for the use of the BW request header and GMSH for specific types of services. Figure 6–5 shows the two GMSHs.

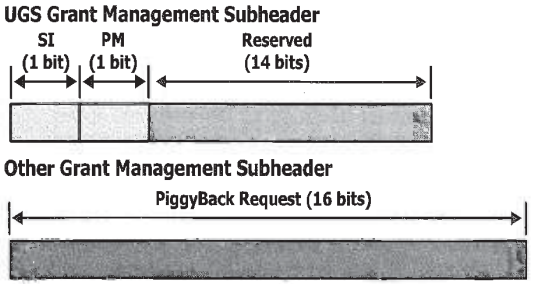


Figure 6–5: GMSHs

In the QoS section of Chapter 7, we describe the different types of per-connection QoS supported by IEEE 802.16. One such QoS type is called *unsolicited grant service* (UGS), in which a connection may be set up with a periodic grant. The UGS GMSH is used by such connections to indicate the status of the current allocation. The slip indicator (SI) bit indicates that the UGS grants are not keeping up with the backlog of the connection queue. The poll-me (PM) bit is used to request a bandwidth poll. The GMSH is always piggybacked on an MPDU of the connection for which additional bandwidth (or signaling) is being requested.

All non-UGS connections use the simple 16-bit GMSH that indicates the number of bytes requested, not including the PHY overhead. Recall that the stand-alone BW request header can be used for aggregate or incremental BW requests. The piggybacked requests are always incremental. Since the size of the GMSH is 2 bytes, it can be used to request up to 65 535 bytes of incremental bandwidth. The stand-alone MAC header allows incremental or aggregate BW requests of up to 524 287 bytes.

Mesh subheader (MSH)

The 2-byte MSH (Figure 6–6), used only by a mesh network, carries the mesh Node ID discussed in Chapter 9.

0085

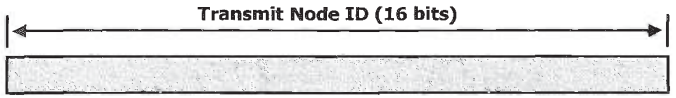


Figure 6–6: MSH

Fast-feedback allocation subheader (FFSH)

The FFSH (Figure 6–7) is used by the BS to request feedback from an AAS SS. The support for AAS functionality is PHY-specific and described in Chapter 13. The allocation offset is defined in units of slots. This is the offset from the beginning of the fast-feedback UL bandwidth allocation of the slot in which the SS must send a fast-feedback message. The allocation applies to the UL subframe of exactly two frames after the frame including the FFSH. The Feedback Type field identifies the type of the feedback, as shown in Table 6–7. Chapter 13 describes these feedback types in detail.

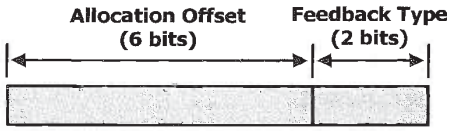


Figure 6–7: FFSH

Table 6–7: Fast-feedback allocation feedback type encodings

Feedback type	Description
00	Fast DL measurement
01	Fast MIMO feedback, antenna #0
10	Fast MIMO feedback, antenna #1
11	MIMO mode and permutation mode feedback

ARQ feedback

ARQ feedback refers to the ARQ control information, such as acknowledgments from a receiver to the transmitter. Since this control information flows in the direction opposite to that of the actual ARQ data flow, efficient mechanisms are required to transport the feedback as a piggybacked payload or as a separate management message. In order to send a separate ARQ feedback management message, especially in the UL, the BS has to make specific allocations for the feedback, or the SS has to steal bandwidth from another allocation. An alternate approach is to piggyback on another connection, which may effectively steal some of its bandwidth.

IEEE Std 802.16 allows such piggybacking of ARQ feedback on any MPDU, subject to system policies. This is similar to the way in which subheaders are transported with MPDUs. However, ARQ feedback is a special payload, not a MAC subheader. The format of the ARQ special payload and rules for piggybacking are described in later sections.

DATA AND MANAGEMENT PDU CONSTRUCTION

Simple MPDU

A simple MPDU consists of the generic MAC header and optionally the payload and CRC. Although CRC is optional, it is a good practice to include it in real-world implementations, unless perhaps in the case of an LOS system with a very low bit error rate (BER). The IEEE 802.16 standard mandates the implementation of CRC if the PHY type is WirelessMAN-SCa, WirelessMAN-OFDM, or WirelessMAN-OFDMA. Figure 6–8 shows the simple MPDU structure.

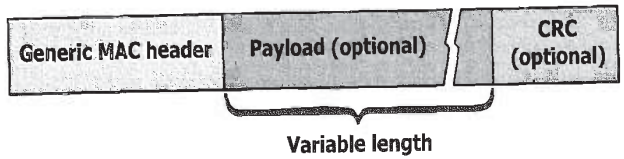


Figure 6–8: MPDU structure

The MPDU format is flexible enough to carry both small and large payloads with minimal overhead. With the appropriate choice of packing and fragmentation and an intelligent scheduler, the overall overhead can be low. When subheaders and special payloads are present, all per-PDU subheaders are placed just after the generic MAC header. The optional subheaders, if present, are considered part of the payload and hence protected by the same CRC that protects the rest of the payload. The format of the MPDU, with subheaders and ARQ special payloads, is shown in Figure 6–9. As shown in the figure, all per-PDU subheaders precede per-SDU subheaders.

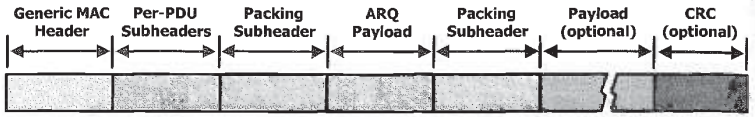


Figure 6–9: MPDU with subheaders

Subheader ordering

As discussed in earlier sections, the generic MAC header indicates the presence or absence of subheaders. However, in order to properly decode MPDUs and subheaders, an ordering of the subheaders and special payloads is defined. The following rules define the use of subheaders in UL and DL:

- ***PSH and FSH:*** The PSH and FSH are mutually exclusive. The FSH is used only to transport a single fragment. The PSH contains the necessary information to support both packing and fragmentation. Therefore, the PSH is used to transport more than one full SDU and/or fragment in the same payload.
- ***ARQ feedback payload:*** The ARQ feedback special payload, if present, is always inserted as the first payload (i.e., after the subheaders, if any) when packing is enabled. This is illustrated in Figure 6–9. Since ARQ feedback is a payload, a PSH must precede it, if packing is required.

- ***MSH:*** The MSH, if present, precedes all other subheaders.
- ***FFSH:*** The FFSH will always be the last per-PDU subheader. Therefore, only the PSH can follow this subheader.
- ***GMSH and FSH:*** The GMSH precedes the FSH.

Based on these rules, the ordering of the subheaders and the special payload is defined as shown in Table 6–8.

Table 6–8: Subheader and special payload ordering

Order	Subheader
1	MSH (mesh subheader)
2	GMSH (grant management subheader)
3	FSH (fragmentation subheader)
4	FFSH (fast-feedback allocation subheader)
5	PSH (packing subheader)
6	ARQ special payload

Figure 6–10 illustrates the ordering of the subheaders in UL and DL. Note that some subheaders are relevant in only one of these.

ARQ blocks

It is important to explain the concept of ARQ blocks before we discuss the fragmentation and packing in detail. The ARQ block is the basic unit of transmission or retransmission for an ARQ-enabled connection. An MSDU is logically partitioned into blocks of a fixed length, negotiated during connection setup. When the length of the SDU is not an integer multiple of the negotiated block size, the final block of the SDU is formed using the remaining SDU bytes after the final full block has been determined.

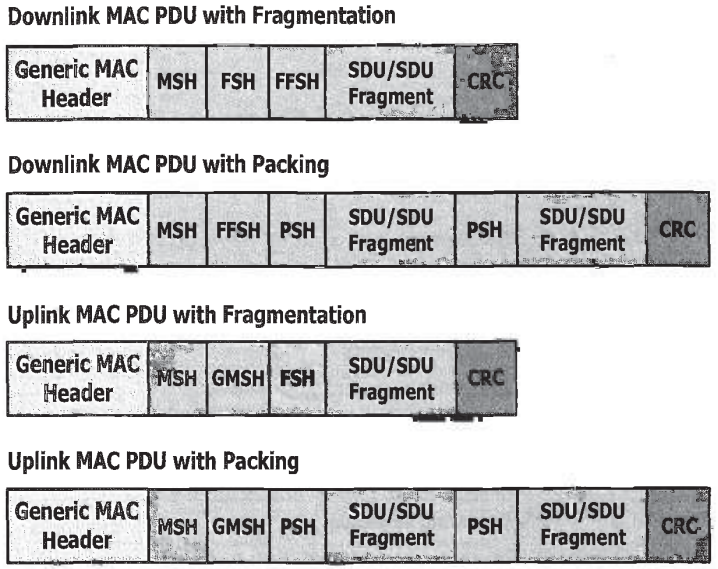


Figure 6–10: Ordering of MAC subheaders

The size of the ARQ block can be of any value from 1 to 2040 bytes. The appropriate choice of the ARQ block size will depend on the ToS carried over the connection, the underlying PHY properties, and the expected feedback delay of the ARQ protocol, among other factors.

Each ARQ block of an SDU is assigned a sequence number called the *block sequence number* (BSN). Fragmentation is allowed to occur only on ARQ block boundaries. Therefore, an ARQ fragment is formed with an integral number of contiguous ARQ blocks. The BSN value carried in the FSH is the BSN for the first ARQ block appearing in the segment.

Figure 6–11 illustrates the concept of ARQ blocks with two consecutive SDUs of the same connection. SDU #1 consists of seven blocks starting from BSN 5. The seventh block, with BSN 11, is of a size smaller than the negotiated ARQ block size. Similarly, SDU #2 consists of five ARQ blocks,

from BSN 12 through BSN 16, where the last ARQ block is smaller than the rest. Both the SDUs are divided into two fragments each: <Frag /SDU#1> (3 blocks, BSN = 5), <Frag 1/SDU#1> (4 blocks, BSN = 8), <Frag 0/SDU#2> (3 blocks, BSN = 12) and <Frag 1/SDU#2> (2 blocks, BSN = 15).

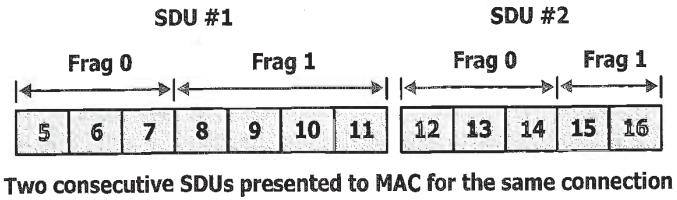


Figure 6–11: ARQ blocks

Fragmentation

The fragmentation function is negotiated on a per-connection basis during connection setup. There are two choices of FSH for non-ARQ connections. The 11-bit FSN is used by all management connections that support fragmentation. The support for the 3-bit FSN is optional in IEEE Std 802.16. If a specific connection has fragmentation enabled, all SDU fragments on that connection are preceded by a FSH. The FSN is incremented by 1 (modulo 8 or modulo 2048) for each SDU fragment and unfragmented SDU.

The fragment size may be chosen dynamically, based on the real-time conditions and other system requirements. This is often called *dynamic fragmentation*. If the regular FSH is used, up to eight fragments are supported per SDU to avoid ambiguity. However, with the extended FSH or when ARQ is enabled, an SDU may be fragmented into any number of fragments to support QoS and other system requirements. There is no restriction of the size of the fragment.

As described in the previous subsection on ARQ blocks, fragmentation works differently for an ARQ connection. For both ARQ and non-ARQ connections, the FSH is of the same size, but the sequence number fields are interpreted

0088

differently. On connections with no ARQ, fragmentation can be done on any byte boundary, and the FSN counter is incremented for every fragment. For ARQ-enabled connections, fragmentation occurs on ARQ block boundaries, and the BSN counter is incremented for every ARQ block.

Fragmentation without ARQ

As soon as a fragment is created, the next valid sequence number is assigned to it. This FSN is carried in the FSH, along with the FC field indicating whether a fragment is the first, an intermediate, or the last fragment from an MSDU. It is also possible to indicate that the "fragment" is a full MSDU (i.e., FC = 00), mostly for convenience of implementation. Although there is no requirement to include a FSH in an MPDU containing a full MSDU, the FSN counter must be incremented even if the FSH is omitted. Figure 6–12 shows the structure of an MPDU with fragmentation.



Figure 6–12: MPDU with fragmentation

The reassembly process for non-ARQ connections in the receiver is fairly straightforward, since fragments are always transmitted in order and can be transmitted only once. If the receiver discovers that a fragment has been lost, it discards any fragments waiting to be reassembled and all fragments until it encounters (a) a fragment indicating it is the first or (b) an MPDU containing a full MSDU. Therefore, the receiver has to buffer at most one full MPDU per connection.

Fragmentation with ARQ

Each MPDU sent on an ARQ-enabled connection must include either the ARQ FSH or the ARQ PSH, even if fragmentation is not enabled, as the BSN must be carried to the receiver. If only fragmentation is used, the MPDU

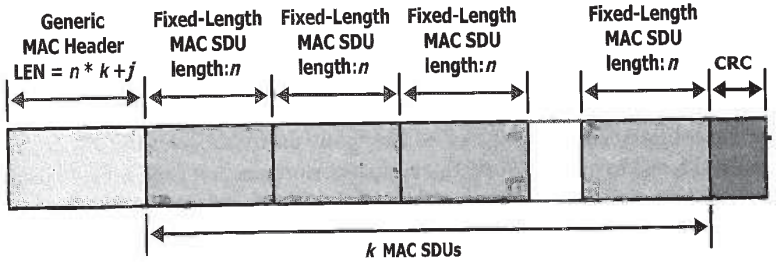
structure is similar to the one shown in Figure 6–12. The only difference is the type of FSH used for ARQ connections.

Packing

There are three types of packing supported in IEEE 802.16: packing of fixed-length SDUs, packing of variable-size SDUs, and packing of ARQ connections.

Packing of fixed-length SDUs

Fixed-size SDUs, such as ATM cells, can be packed and transported on a connection without using PSHs. During connection setup, the SDU parameters can be negotiated for fixed-size packing with no fragmentation. This allows the fixed-size SDUs to be packed efficiently without the PSH overhead. The number of SDUs per MPDU can still be varied dynamically based on QoS and other requirements. The Length field in the generic MAC header implicitly defines the number of packed SDUs. Figure 6–13 shows an MPDU packed with fixed-size SDUs.



Note: j is the length of the MAC Header plus the CRC, in bytes.

Figure 6–13: Packing of fixed-length SDUs

0089

Packing of variable-length SDUs

Variable-length SDUs can be packed using one of the PSHs. Although all PSHs have all the necessary fields to simultaneously support packing and fragmentation, packing can be used without fragmentation. The regular PSH is sufficient to support packing without fragmentation. A packed MPDU, with each SDU preceded by a PSH, is shown in Figure 6–14.

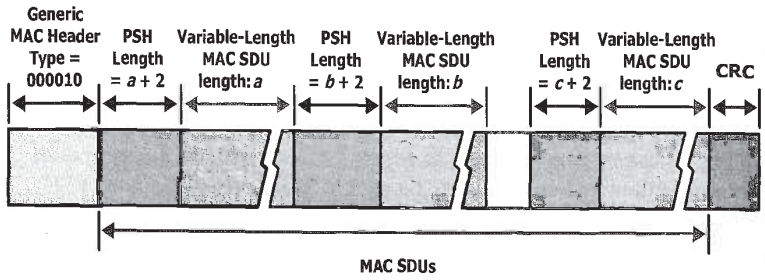


Figure 6–14: Packing of variable-length SDUs

Packing with fragmentation

As discussed earlier, the PSH is a superset of the FSH. Both can be simultaneously supported on the same connection. This combination provides the utmost flexibility to the MAC to efficiently utilize the over-the-air bandwidth and to maximize overall system performance. It is possible to pack multiple consecutive fragments of the same SDU into the same MPDU. However, such fragmentation and packing can only increase the overhead unnecessarily, with no real benefit. Therefore, typically an MPDU with packing and fragmentation will have at most two fragments of two different SDUs and zero or more unfragmented SDUs.

Figure 6–15 shows an example of a packed MPDU with two fragments and some unfragmented SDUs. The first packed payload is the last fragment of an SDU and the last payload is the first fragment of another SDU. Another example of a packed MPDU with one fragment and some unfragmented SDUs is shown in Figure 6–16.

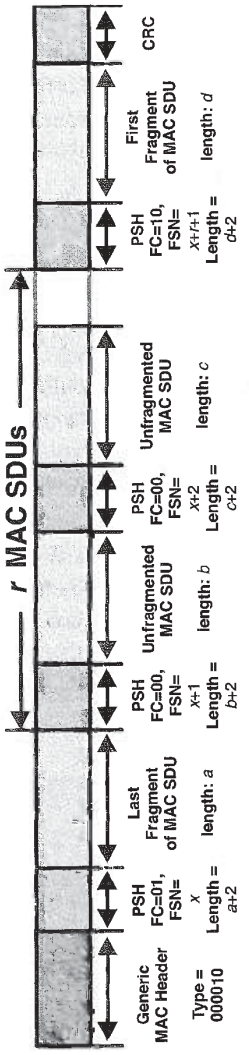


Figure 6–15: Packed MPDU with two fragments

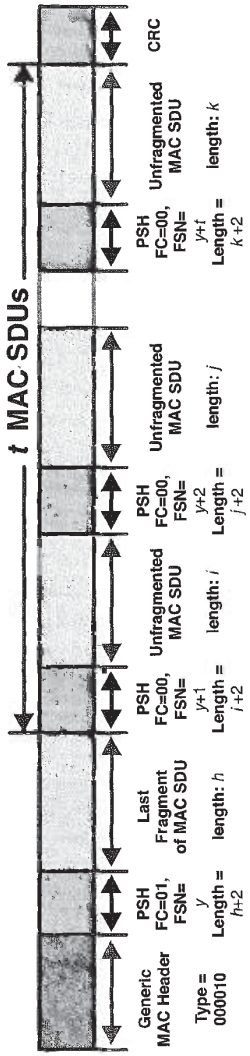


Figure 6–16: Packed MPDU with one fragment

0090

Packing and ARQ

An ARQ connection must use either a FSH or a PSH, as the ARQ algorithm requires the BSN to keep track of ARQ blocks. If ARQ is used with no fragmentation or packing, the ARQ FSH is used to carry the BSN. The same subheader can be used to add fragmentation function to an ARQ connection without packing. If the ARQ connection requires packing (with or without fragmentation), the ARQ PSH must be used.

Packing of ARQ payload

The ARQ feedback payload consists of one or more ARQ Feedback information elements (IEs), the format and contents of which will be covered elsewhere in this book. When the ARQ feedback is piggybacked onto another MPDU with packing enabled, it shall always be the first packed payload. The FC and FSN fields are irrelevant to the ARQ feedback and are set to zero. Since the length of the ARQ feedback payload is variable, the Length field in the PSH is used to correctly decode the ARQ payload from the packed MPDU.

Although only the Length field of the PSH is of significance for packing ARQ feedback payload, all PSHs used within an MPDU should be of the same type (i.e., regular, extended, or ARQ), as negotiated during the connection setup. The ARQ feedback payload can be piggybacked on a connection that may or may not have ARQ enabled. Figure 6–17 shows an example MPDU with an ARQ feedback payload.



Figure 6–17: ARQ feedback and packing

Concatenation

Further efficiency can be achieved by reducing the per-PDU PHY overhead through concatenation. This is achieved by transmitting multiple MPDUs in a single PHY burst. Although concatenation is not a MAC feature, it is covered in this section because it is related to other efficiency functions, such as packing and fragmentation. Concatenation is allowed in both the UL and DL.

For some of the IEEE 802.16 PHY specifications, an intermediate sublayer, called the *transmission CS*, is specified to efficiently support concatenation. This sublayer inserts a pointer byte at the beginning of a payload to help the receiver separate the MPDUs in a burst. This can be used to correctly receive some of the concatenated MPDUs in a PHY burst even if part of the PHY burst was received incorrectly (e.g., if a part of the burst was corrupted). The transmission CS is mandatory for the WirelessMAN-SC PHY and optional for the WirelessMAN-OFDM PHY.

Figure 6–18 shows the concept of concatenation for an UL burst transmission. As shown in the figure, any type of MPDU or message may be concatenated into a single PHY burst. The figure shows two UL bursts, one that concatenates a BW request and a user data MPDU and another that concatenates a MAC management message and user data MPDUs.

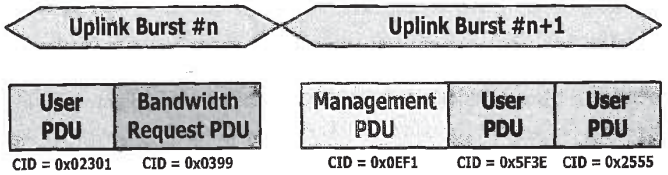


Figure 6–18: Concatenation of MPDUs

MPDU encryption and CRC

The details of the MPDU encryption are covered in Chapter 8. A summary of MPDU encryption and CRC rules are listed here:

- CRC is optional in the IEEE 802.16 MAC, although it is a good practice to include it.
- CRC is mandatory for ARQ connections, as it is used to detect errors in MPDU transmissions.
- If CRC is included, it covers both the MAC header and payload. The generic MAC header is also protected by HCS.
- If encryption is used, the generic MAC header is not encrypted.
- If both CRC and encryption are enabled, CRC is calculated after encryption.

MAC management

The format of the MAC management messages is shown in Figure 6–19. The MAC management messages are carried as (a part of) MPDU payload. All MAC management messages begin with an 8-bit Type field, followed by the message payload. An example management message payload definition is shown in Table 6–9.

The first part of the payload follows no particular set of encoding rules. In the example, it a string of 3 bytes, the two first of which are to be interpreted as the transaction ID and the third one as the confirmation code. The remainder of the message is defined as *TLV-encoded information*, which follows an encoding that closely resembles ASN.1 Basic Encoding Rules. For each parameter encoded as a type-length-value (TLV), the first byte identifies the parameter type, the following byte/bytes[2] gives the length in bytes of the value field, and the actual parameter value follows.



Figure 6–19: Format of the MAC management message

[2] If the length of the Value field is less than 128 bytes, a single byte suffices. For longer parameters, more than 1 byte is used to express the length.

Table 6–9: DSA-RSP message format

Syntax	Size	Notes
DSA-RSP_Message_Format() {		
Management Message Type = 12	8 bits	
Transaction ID	16 bits	
Confirmation Code	8 bits	
TLV Encoded Information	*variable*	TLV-specific
}		

ARQ

The following subsections describe the basic IEEE 802.16 ARQ concepts of block-based retransmissions and bitmap-based ARQ feedback. The actual SR ARQ algorithm, with its operational parameters and state machines, are described in Chapter 7.

ARQ block-based retransmissions

The concept of ARQ blocks and block-based retransmissions is illustrated in Figure 6–20. The figure shows the transmission, loss, and subsequent retransmission of two consecutive SDUs, shown in Figure 6–11, on an ARQ-enabled connection with packing. The number of blocks per SDU is based on the prenegotiated block size. During initial transmission of the SDUs, both are fragmented into two fragments each. The number of fragments per SDU and hence the number of blocks per fragment are dynamically decided based on a variety of factors. For example, the scheduler may limit the fragment size or airtime allocated to a specific connection. Even if the allocated airtime is fixed, the adaptive modulation may change the number of blocks that can be transmitted within the allocated time.

0092

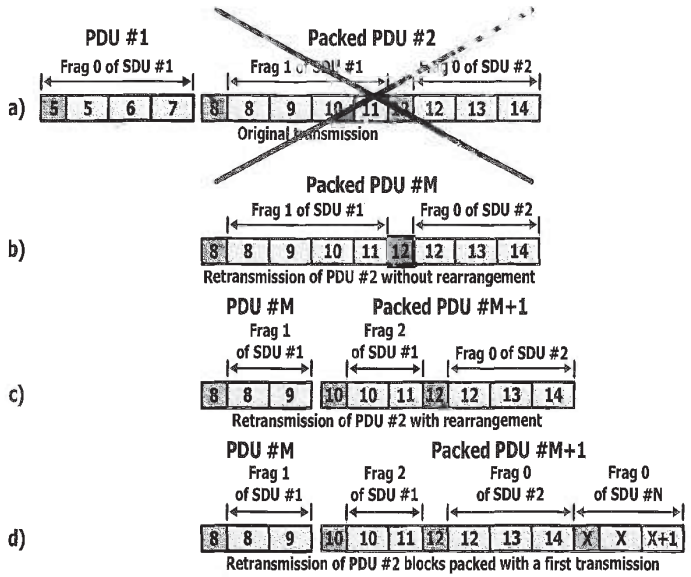


Figure 6–20: ARQ blocks and retransmissions

Figure 6–20 a) shows the construction and transmission of two MPDUs, where PDU #1 consists of fragment #0 of SDU #1, and PDU #2 consists of fragment #1 of SDU #1 and fragment #0 of SDU #2. The transmission of fragment #2 of SDU #2 is not shown in the figure. Assume it is determined (by mechanisms discussed in the next subsection) that PDU #2 did not reach the receiver. The block-based approach offers many possibilities for retransmission of the lost PDU #2. We assumed that, by the time the loss of PDU #2 is detected, fragment #1 of SDU #2 (consisting of blocks 15 and 16) has already been transmitted successfully and that SDU #*N* of the same connection, where *N* is greater than 2, is ready for first transmission.

Out of the many possibilities for the retransmission of PDU #2, three examples are shown in Figure 6–20. The first example, in Figure 6–20 b),

shows the retransmission of PDU#2 with no change or rearrangement as a new PDU #*M* (*M* > 2). The second example, in Figure 6–20 c), shows the retransmission with a rearrangement where the original PDU #2 is split into two MPDUs, *M* and *M*+1. The third example, in Figure 6–20 d), is similar to the second one, except the PDU #*M*+1 consists of the blocks to be retransmitted and the first or original transmission of fragment #0 (with BSN *X*) of SDU #*N*. As shown in Figure 6–20 d), an MPDU on an ARQ connection may contain blocks that are transmitted for the first time as well as those being retransmitted.

It is up to the transmitter's policy to decide whether a set of blocks once transmitted as a single fragment should be retransmitted as a single fragment. Similarly, it is up to the transmitter to decide whether a set of fragments once transmitted in a single MPDU should be retransmitted as a single MPDU. The advantage of keeping the number of blocks per fragment and fragments per MPDU the same between retransmissions is simplicity. On the other hand, the flexibility of refragmentation and rearrangement can significantly improve efficiency and QoS in certain traffic conditions. If fragmentation is not enabled for this connection, the ARQ fragmentation or PSH would still be used with the same block size, as if the fragmentation or packing were enabled. In this case, each "fragment" would contain all the blocks of a single SDU.

ARQ Feedback information element (IE)

Apart from the flexible retransmission, IEEE Std 802.16 defines compact ARQ feedback that is used by the receiver to provide feedback on successfully received and missing blocks. The ARQ Feedback IE contains an ACK Type field and an optional variable-size bitmap indicate positive and/or negative acknowledgments. Four types of acknowledgments are defined to support flexible feedback. Figure 6–21 shows the format of the ARQ Feedback IE.

0093

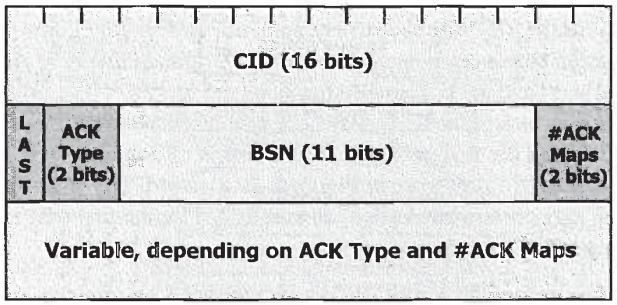


Figure 6–21: ARQ Feedback IE

The fields of the ARQ Feedback IE are described below:

- ***CID:*** The CID of the connection for which this feedback is being sent.
- ***Last:*** The Last bit is used to pack multiple ARQ Feedback IEs to create an ARQ feedback payload. The Last bit is set to zero for all ARQ Feedback IEs, except the last IE in the payload.
- ***ACK Type:*** This field indicates the type of acknowledgment being transported in this feedback IE. Table 6–10 shows the acknowledgment types and their descriptions.
- ***BSN:*** The BSN is interpreted differently based on the acknowledgment type as described in Table 6–10.
- ***# ACK MAPs:*** The value of this field plus one represents the total number of 16-bit acknowledgment bitmaps included in this feedback IE. This field is not valid for acknowledgment type 0b01 and, therefore, is set to zero.
- ***ACK MAPs:*** The number of acknowledgment MAPs is variable, and the acknowledgment type 0b01 does not have an associated acknowledgment MAP. The acknowledgment MAPs are treated as a bit stream, where the MSB is the MSB of the first acknowledgment MAP and the least significant bit (LSB) is the LSB of the last acknowledgment MAP.

Table 6–10: Acknowledgment types

Value	Type	Description
00	Selective	The BSN carried within this IE corresponds to the MSB of the first 16-bit acknowledgment bitmap. Each subsequent bit corresponds to the BSN corresponding to the previous bit plus one. No assumptions should be made about the status of any of the BSNs prior to the first BSN or the last BSN represented by the acknowledgment MAPs. If the bit is set, it indicates that the receiver has successfully received the corresponding ARQ block.
01	Cumulative	The ARQ block indicated by the BSN and all ARQ blocks prior to this within the transmission window have been successfully received. No ARQ MAP is included with this acknowledgment type.
10	Cumulative with selective	This is the combination of the previous two acknowledgment types (0b00 and 0b01). Combines the functionality of types 0x0 and 0x1. The first bit of the first acknowledgment MAP, which corresponds to the included BSN, is always set to 1, indicating cumulative acknowledgment for the BSN.
11	Cumulative with block sequence	This acknowledgment type combines the functionality of the cumulative acknowledgment type 0b10 with the ability to acknowledge a sequence of ARQ blocks in a compact format. If a set of ARQ blocks with consecutive BSN values all have the same reception status (i.e., either positive or negative), a single bit indicating the status, together with a sequence length, is used to indicate the status of the whole sequence of BSNs.

Since all the bits in a selective acknowledgment MAP must refer to an ARQ block that was successfully received or an ARQ block that was presumed lost or received incorrectly, it is important to choose the right BSN for the Selective ACK Feedback IE. The receiver must assign a BSN so that all bits in a MAP either indicate a positive acknowledgment (ACK) or a negative acknowledgment (NACK). One way to achieve this is to choose a lower BSN so that all bits are valid. For example, if only 12 ARQ status bits are available (either ACK or NACK) starting from BSN *N*, the receiver may send a selective acknowledgment with a single 16-bit MAP and a BSN of *N - 4*, where all the first 4 bits will be set to one.

The format of the block sequence acknowledgment MAP is shown in Figure 6–22. There are two block sequence formats. The first (and most significant) bit of the block sequence acknowledgment MAP indicates the format. The first format includes a 2-bit sequence acknowledgment MAP, where the bits are interpreted similarly to the selective acknowledgment type, except that each bit corresponds to a BSN sequence instead of a single BSN. Since this is also a cumulative acknowledgment, the BSN carried within this ACK Feedback IE indicates that the ARQ block with this BSN and all ARQ blocks with lesser BSN values within the transmission window have been successfully received. However, the first bit of the sequence acknowledgment MAP corresponds to the first BSN of a sequence that is being selectively acknowledged, which is the BSN carried in the feedback IE plus one. Therefore, the first bit may be a negative acknowledgment, unlike the acknowledgment type 0b10, where the first bit must always be set to one, as it corresponds to the cumulative acknowledgment BSN. The length of the each BSN sequence is indicated in the two 6-bit sequence length fields that follow. The first format can be used to indicate the selective acknowledgment status of up to 126 ARQ blocks.

The second format shown in Figure 6–22 is very similar, except that it has a 3-bit sequence acknowledgment MAP, where the length of each sequence is indicated by the 4-bit sequence length fields. Therefore, a total of up to 45 ARQ blocks can be selectively acknowledged using this 16-bit compact format.

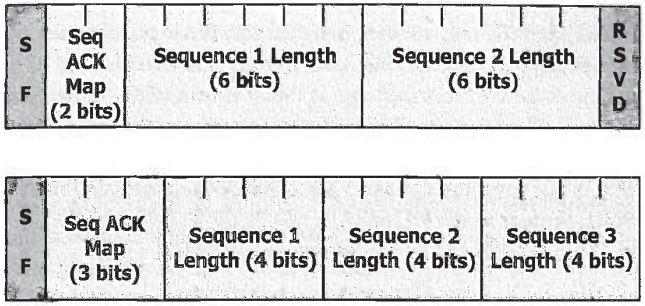


Figure 6–22: Block sequence acknowledgment MAP formats

ARQ feedback payload

The ARQ feedback information can be either sent as a stand-alone MAC management message on the appropriate basic management connection or piggybacked as a special payload on an existing connection. ARQ feedback sent in a basic management connection cannot be fragmented. When transported in a management connection, more than one ARQ Feedback IE may be sent, and the Last bit is used to properly decode the contents.

HYBRID AUTOMATIC REPEAT REQUEST (HARQ)

The HARQ is an optional part of the MAC. HARQ is supported only for the OFDMA PHY. HARQ, along with associated parameters, is negotiated during initialization of the SS. HARQ defines four basic concepts to efficiently support retransmissions:

- Compact MAP IEs that define access information for both UL and DL
- HARQ packet construction
- Reduced connection identifiers (RCIDs)
- HARQ UL/DL acknowledgments

0095

Compact MAP IE

The Compact MAP IE may include both DL/UL IEs that define the access information for the DL and UL burst of HARQ–enabled SS. One distinguishing feature of this message is that it is sent without the generic MAC header, unlike all other MAPs, in order to save a few bytes of overhead. This is accomplished by setting the first 3 bits of the HARQ MAP message to a value that is not valid for a generic MAC header encapsulated management or data PDU. Similar to the DL and UL MAPs for TDM and TDMA operation, these HARQ MAPs are broadcast by the BS.

A BS may broadcast multiple HARQ compact MAP messages using multiple burst profiles (i.e., modulation and coding rate). The HARQ compact MAP always follows the DCD and UCD messages, if present. Otherwise, HARQ compact MAPs are sent after other DL/UL MAP messages. When multiple HARQ compact MAP bursts are present, they are ordered based on the modulation and coding used, starting from the lowest rate to the highest. A HARQ MAP Pointer IE is used by the BS to indicate to the SS the number of slots allocated to the OFDMA burst containing HARQ MAP messages.

HARQ Control IE

HARQ Control IEs that are part of the Compact MAP IEs carry the necessary encoding/decoding information of the HARQ–enabled DL/UL bursts. This information is used to properly decode the HARQ subpackets. The information contained in the HARQ Control IE includes the following:

- A 1-bit HARQ sequence number, toggled between 1 and 0
- A 2-bit SPID that identifies the subpacket as described in the following subsection
- A 4-bit HARQ channel ID that identifies the HARQ channel

Since HARQ is a stop-and-wait ARQ, a single bit sequence number is sufficient for proper operation. However, more than one HARQ channels may be defined per terminal, where each connection may have an encoded packet pending. The HARQ Control IE may also be used by the BS to request a retransmission of a subpacket.

Construction of HARQ packets

The construction of an HARQ packet involves the concatenation of one or more MPDUs, as shown in Figure 6–23. A CRC is added to this concatenated packet to form an HARQ PHY burst.

After the construction, an HARQ packet is encoded according to the PHY specification, and four subpackets are generated from the encoded result. The subpackets are identified by an SPID. The encoding and generation of four subpackets introduce redundancy among the subpackets. Therefore, the subpackets can be transmitted one at a time, in sequence, and the receiver can correctly decode the original encoder packet without having to receive all four subpackets.

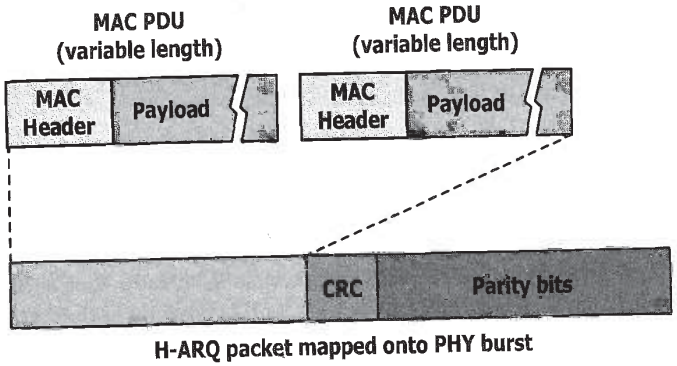


Figure 6–23: Construction of HARQ packet

Reduced connection identifier (RCID)

Figure 6–24 shows the format of a RCID. An OFDMA BS may use RCID instead of a Basic CID or multicast CID to reduce the size of MAP messages, including HARQ MAP messages. There are three types of RCIDs defined in IEEE Std 802.16: RCID-11, RCID-7, and RCID-3. The type of RCID used in a specific system is determined by BS considering the range of Basic CIDs of the SSs connected with the BS. The RCID type to be used is specified in the

RCID_Type field of the Format Configuration IE that is used in the OFDMA PHY to select format configuration.

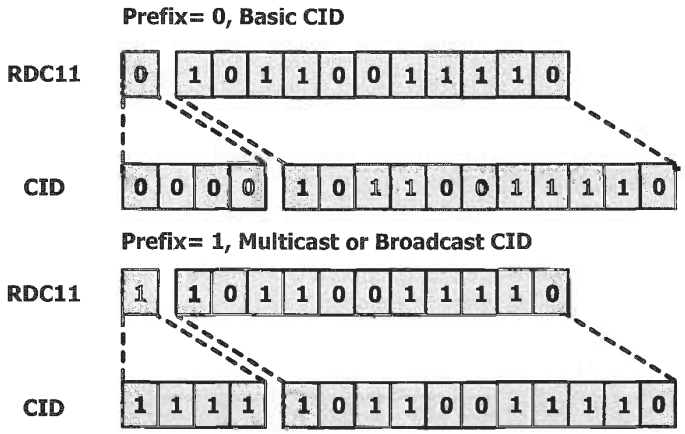


Figure 6–24: RCID decoding

The RCID is carried in a RCID_IE, whose fields are interpreted based on the RCID_Type field (from the Format Configuration IE) and the Prefix field in the RCID_IE itself, if applicable. When the RCID_Type field value is nonzero, the RCID_IE is composed of 1 bit of prefix and *n*-bits of LSB of the SS CID. The prefix is set to one for the Broadcast CID or Multicast Polling CID and set to zero for the Basic CID. If the RCID_Type field value is zero, then the RCID_IE consists of a regular 16-bit CID. Therefore, an RCID_IE can be either 16 bits, 12 bits (RCID-11), 8 bits (RCID-7), or just 4 bits (RCID-3). Table 6–11 shows the interpretation of the RCID_IE fields, where the [R]CID length and RCID_IE length are specified in bits.

Table 6–11: RCID field interpretation

RCID_Type	Prefix	[R]CID length (bit)	RCID_IE length (bit)	Comments
0	NA	16	16	This is the same as normal CID
NA	1	11	12	11 LSB of Multicast, AAS, or Broadcast CIDs
1	0	11	12	11 LSB of Basic CID
2	0	7	8	7 LSB of Basic CID
3	0	3	4	3 LSB of Basic CID

The RCID cannot be used in place of the Transport CID, Primary Management CID, or Secondary Management CID.[3] For the purpose of per-SS HARQ, an SS is always identified by its Basic CID (or a corresponding RCID for the Basic CID).

HARQ acknowledgments

The HARQ acknowledgments are handled differently in the UL and DL. The BS uses a Compact MAP IE to send HARQ acknowledgment bitmap information for the HARQ UL bursts from the SS. The BS specifies a "frame offset" that indicates the offset between the UL HARQ burst and the corresponding HARQ bitmap. This frame offset f is communicated in the HARQ_ACK_ Delay_for_UL_Burst field of the DCD messages. The SS uses the frame offset to correlate acknowledgment bitmaps to the HARQ bursts. For an HARQ UL burst in the i^{th} frame corresponding to the j^{th} HARQ burst in the UL MAP, the BS uses the j^{th} bit (LSB) of the HARQ bitmap sent in the $(i+f)^{th}$ frame to indicate the status. The DL Compact MAP IE (type = 5) that

[3] For other uses of RCID in OFDMA, an exception to this rule is made for the Multicast Polling RCID in the DL. If a DL CID with prefix 1 and RCID 11 is in the range of the Multicast Polling CID (0xFF00–0xFFFD), then the DL CID is interpreted as a DL Transport CID by subtracting 0xFF (0xFE01–0xFEFE).

0097

carries the HARQ bitmap consists of a Length field that indicates the length of the bitmap and the actual bitmap.

The DL HARQ bursts are acknowledged in the HARQ region allocated by the BS in a UL Compact MAP IE. The OFDMA subchannels in the HARQ region are divided into two half-subchannels, where the $(2n)^{th}$ half-subchannel is the first half-subchannel and the $(2n+1)^{th}$ half-subchannel is the second half-subchannel of the n^{th} subchannel. An HARQ DL burst at the i^{th} frame is acknowledged by an SS through the half-subchannel in the HARQ region at the $(i+f)^{th}$ frame, where the frame offset f is defined by the HARQ_ACK_Delay_for_DL_Burst field in the UCD message. The order of DL HARQ burst defined in the HARQ MAP determines the half-subchannel offset in the HARQ region. For an HARQ–enabled burst at the i^{th} frame that is the n^{th} HARQ–enabled burst in the HARQ MAP, the SS transmits the HARQ acknowledgment at the n^{th} half-subchannel in the HARQ region that is allocated by the BS at the $(i+f)^{th}$ frame.

Chapter 7 MAC operation

Radio control, QoS, and ARQ

NETWORK ENTRY AND INITIALIZATION

Once the SS has powered up, it begins the network entry and initialization process. After completing the steps of the process, the SS has all the addresses and parameters it needs to communicate with the rest of the network.

Scanning and synchronization to the DL

Before the SS can do anything else, it has to first find a BS transmitting a signal that it is capable of decoding and understanding. To accomplish this, the SS consults the frequency lists with which it shipped, chooses one of the channels mentioned in the list, and starts to search for the periodically occurring frame preamble. If the SS finds the preamble, it tries to determine the DL transmission parameters by looking at the frame control header (FCH) and the DCD message, which is transmitted periodically but not in every frame. Also, the SS learns the specified UL parameters from the UCD messages. Afterwards, it has sufficient knowledge to transmit on the UL, and initial ranging may begin.

Initial ranging

Initial ranging accomplishes the following:

- The timing advance of SS transmissions is adjusted to make the SS appear collocated with the BS.
- The transmission power of the SS is adjusted for optimal reception at the BS.
- The SS is allocated its Basic and Primary Management CIDs.
- In some systems, the frequency of SS UL transmissions is adjusted.

Initial ranging occurs in a part of the frame that is set aside for this purpose and partitioned into initial ranging slots. The slots need to be large enough to accommodate the ranging transmission and twice the propagation delay to the edge of the cell, since the distance between SS and BS is not known at this time. The exact details of the initial ranging procedure depend on the underlying PHY, but in short the process goes as follows: First, the BS signals the available initial ranging slots in the UL MAP. After determining the location in time of the initial ranging slots, the SS randomly selects a slot and sends an RNG-REQ (ranging request) message. If this is received correctly, the BS replies with a message specifying the power, timing, and possibly frequency adjustments the SS needs to make prior to its next transmission. This message can also instruct the SS to repeat the initial ranging process. However, the BS would normally want to instead fine-tune the SS transmission parameters using one of the maintenance ranging procedures, as they are less resource-consuming than initial ranging.

Now, there is always the possibility that two terminals may send their RNG-REQs in the same slot, in which case the BS may be unable to decode either transmission. Also, it might be that the SS transmission fails because its power is too low (noting that the SS will initially use low power so it does not disrupt transmission in neighboring cells). In any case, if the SS does not hear from the BS within a certain interval, it will increase the power by a notch and try again after waiting for a randomly chosen number of initial ranging slots. In addition, for some PHYs, if the BS can detect energy but fails to decode the initial ranging message, it can signal for the SS to retransmit and make adjustments, identifying the SS in the RNG-RSP (ranging response) message by the transmit slot.

SS basic capability negotiation

Once the SS timing and transmit power have been adjusted, the SS needs to tell the BS which optional functionality it supports. Also, the BS needs to tell the SS which of those options the SS is allowed to use. The protocol by which the support and use of options are agreed upon is called *SS basic capability negotiation*. As this name suggests, this basic negotiation covers not all options, but only those having to do with fundamental MAC and PHY features, such as maximum transmit power, modulation schemes, FEC codes,

and support of MAC bandwidth allocation schemes. The idea with exchanging these parameters at an early stage is to make the rest of the network entry process more resource-efficient.

The protocol itself is a straightforward request-grant exchange, as shown in Figure 7–1. The timer T18 has a default value of 50 ms and should be significantly less than the registration timeout (i.e., the timeout at the BS for the maximum time between a RNG-RSP to an SS upon successful ranging and the receipt of an SBC-REQ (SS basic capability request) message from the same SS). In the request message (SBC-REQ), the SS will list all the "basic capabilities" it supports. The BS inspects the list of capabilities and, in the SBC-RSP (SS basic capability response) message, informs the SS of those it does not support or does not want to use.

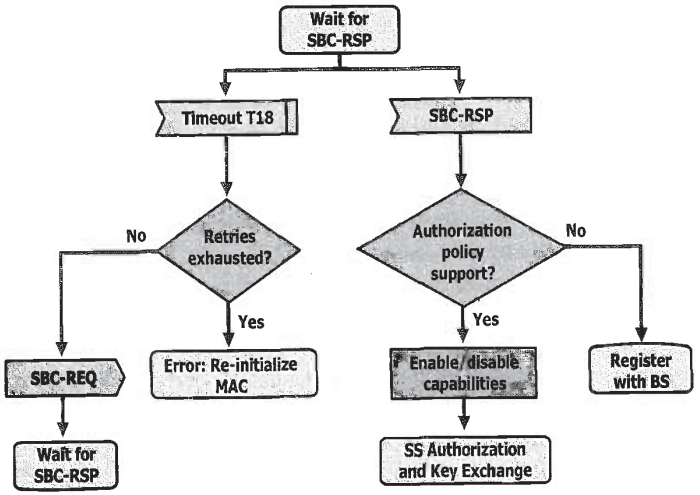


Figure 7–1: SS basic capability negotiation

0099

Note that some of the most fundamental parameters of operation are not subject to negotiation with this protocol. For example, with different PHY system profiles, establishing communication would be challenging or impossible.

Authorization, security association (SA) establishment, and key exchange

At this stage of the network entry process, the SS should be capable of communicating efficiently with the BS. The BS, however, does not yet know with certainty the identity of the SS with which it is dealing. The SS has up to this stage provided its IEEE MAC address in the RNG-REQ, but has offered no credentials verifying its identity. The authorization phase establishes the identity of the SS, the authorization key (AK), and the list of SAs that the SS is authorized to use. To achieve this, well-known public key cryptographic methods are used. For a treatise on these, consult [B45].

The authorization protocol starts with the SS sending the authentication information privacy key management (PKM) message. All PKM messages are encapsulated in either the PKM-REQ (SS to BS) or the PKM-RSP (BS to SS) management message. The authentication information message is strictly informational, but is included to provide a mechanism by which a BS can learn the X.509 certificate of the SS manufacturer in a standard way. However, the IEEE 802.16 standard says nothing about how the BS can authenticate the certificate. It is envisaged that the industry will set up the needed infrastructure for this authentication if this method is to be used in practice. Meanwhile, the assumption is that the manufacturer certificate ends up, by some secure means, in the BS. Immediately after sending the authentication information message, the SS sends the Authorization Request message encapsulated in a PKM-REQ management message containing the X.509 certificate of the SS together with a list of the SS security capabilities. The BS proceeds to authenticate the certificate by checking the signature of the manufacturer. Assuming the certificate is in order, the BS generates the AK and encrypts it using the public key of the SS contained in the X.509 certificate. Additionally, the BS determines which SAs the SS is entitled to use. The AK, the key lifetime, and the list of SAs along with their parameters are sent by the BS to SS in the Authorization Reply message. Also, the BS has

the option to send any PKM protocol parameters that are different from the default values given in the IEEE 802.16 standard.

After receiving the Authorization Reply message and decrypting the AK, the SS becomes authorized to operate normally in the network. However, it still lacks the capability to transmit anything besides MAC management messages. Before becoming fully operational, the SS still needs to (a) acquire a key for each SA given by the BS using the protocol given in the SA parameter perform registration, (b) possibly acquire an IP address, and (c) establish the MAC connections. Each SA's keys, called *traffic encryption keys* (TEKs), are established by the SS sending a digitally signed Key Request PKM message and the BS returning the encrypted key in a Key Reply message, also signed with a key derived from the AK. The algorithm used in the encryption of the TEK is triple data encryption standard (3DES), AES, or RSA, depending on algorithm support of the SS and the parameters of the SA.

Registration

During the registration phase, the SS and BS negotiate additional operational parameters of the MAC. Also, the SS informs the BS whether it will be part of the managed network. If the SS is to be managed, the bidirectional secondary management connection is established between the SS and BS. The registration message exchange also offers the possibility to negotiate the version of IP and the QoS parameters for the secondary management connection. The registration process is shown in Figure 7–2. The SS may wait for up to 3 s (timer T6 in Figure 7–2) for the registration response before resending the request or reinitializing the MAC.

Establish IP connectivity

A managed SS also needs to acquire an IP address (using DHCP), pull down a configuration file (using TFTP), and establish the time of day (using the Internet Time Protocol). Afterwards, the SS sends a message to report that the tasks were successfully completed.