CHAPTER

IEEE 802.16m Reference Model and Protocol Structure

3

INTRODUCTION

The IEEE 802.16-2009 standard defines a generic reference model where major functional blocks (i.e., physical layer, security sub-layer, MAC common part sub-layer, and service specific convergence sub-layer) and their interfaces, the premises of IEEE 802.16 entity, and a general network control and management system are specified. The IEEE 802.16m has modified this reference model by further classifying the MAC common part sub-layer functions into two functional groups, resulting in a more structured approach to characterizing the data link layer functions and their interoperation.

The earlier revisions and/or amendments of the IEEE 802.16 standard did not explicitly define any detailed protocol structure; rather, the functional elements in the specification were implicitly classified as convergence sub-layer, MAC common part sub-layer, security sub-layer, and physical layer. While each of these layers and/or sub-layers comprises constituent functions and protocols, no perspective was provided on how various components were interconnected and interoperated from a system standpoint. In fact, the IEEE 802.16 standards have never been developed with a system engineering approach; rather, they specify components and building blocks that can be integrated (obviously various combinations are potentially possible) to build a working and performing system. An example is the mobile WiMAX system profiles [1], where a specific set of IEEE 802.16-2009 features were selected to form a mobile broadband wireless access system. In an attempt to improve the clarity of the previous IEEE 802.16 standards and to take a systematic approach in development of the advanced air interface, IEEE 802.16m has defined a protocol structure and the functional components are classified into different layers and sub-layers, as well as differentiated based on data-plane or control-plane categories.

The protocols and functional elements defined by the IEEE 802.16 standard correspond to the physical and data link layers of the Open System Interconnection (OSI) seven-layer network reference model as shown in Figure 3-1.

In the context of protocol structure, we will frequently use the terms “service” and “protocol.” It must be noted that services and protocols are distinct concepts. A service is a set of primitives or operations that a layer provides to the layer(s) with which it is interfaced [2]. The service defines what operations a layer performs without specifying how the operations are implemented. It is further related to the interface between two adjacent layers. A protocol, in contrast, is a set of rules presiding over the format and interpretation of the information/messages that are exchanged by peer entities within a layer. The entities use protocols to implement their service definitions. Thus, a protocol is related to the implementation of a service.

Mobile WiMAX. DOI: 10.1016/B978-0-12-374964-2.10003-7

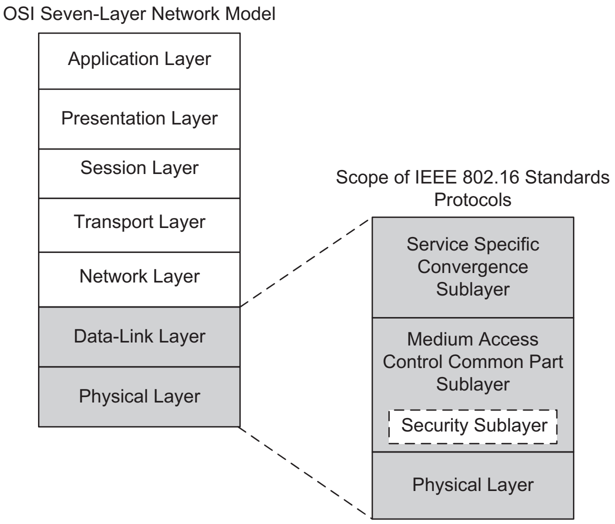

OSI Seven-Layer Network Model
Application Layer
Presentation Layer
Session Layer
Transport Layer
Network Layer
Data-Link Layer
Physical Layer
Scope of IEEE 802.16 Standards Protocols
Service Specific Convergence Sublayer
Medium Access Control Common Part Sublayer
Security Sublayer
Physical Layer

FIGURE 3-1

The mapping of IEEE 802.16 protocol layers to an OSI seven-layer network model

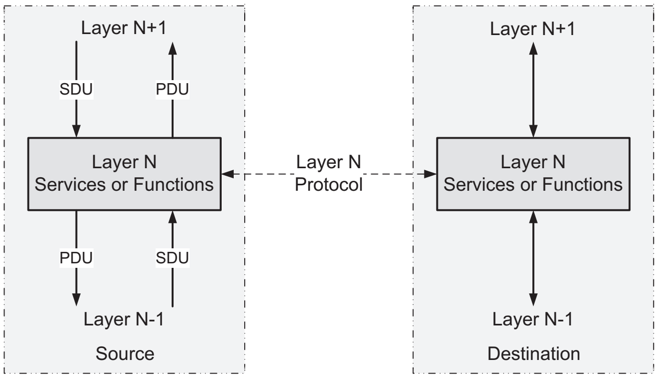


FIGURE 3-2

An illustration of service, protocol, PDU, and SDU concepts [2]

As shown in Figure 3-2, a Protocol Data Unit (PDU) is a packet exchange between peer entities of the same protocol layer located at the source and destination. On the downward direction, the PDU is the data unit generated for the next lower layer. On the upward direction, it is the data unit received from the previous lower layer. A Service Data Unit (SDU), on the other hand, is a data unit exchanged between two adjacent protocol layers. On the downward direction, the SDU is the data unit received from the previous higher layer. On the upward direction, it is the data unit sent to the next higher layer.

This chapter provides a top-down systematic description of IEEE 802.16m reference model and protocol structure, starting at the most general level and working toward details or specifics of the protocol layers, their functional constituents and interconnections. An overview of 3GPP LTE protocol structure is further provided to enable readers to contrast the corresponding protocols and functionalities.

It must be noted that while the IEEE 802.16 standard does define a generic network reference model (or a network abstraction model), the mobile WiMAX systems use the specific network reference model and system architecture that were described in Chapter 2 to achieve interoperability. Therefore, the network reference model and associated components and interfaces described in Section 3.1 are only informative, and they should not be interpreted as normative for implementation and deployment of the IEEE 802.16m systems.

3.1 THE IEEE 802.16m REFERENCE MODEL

Figure 3-3 illustrates the IEEE 802.16 reference model [3]. The data link layer of IEEE 802.16 standard comprises three sub-layers. The service-specific convergence sub-layer (CS) provides any

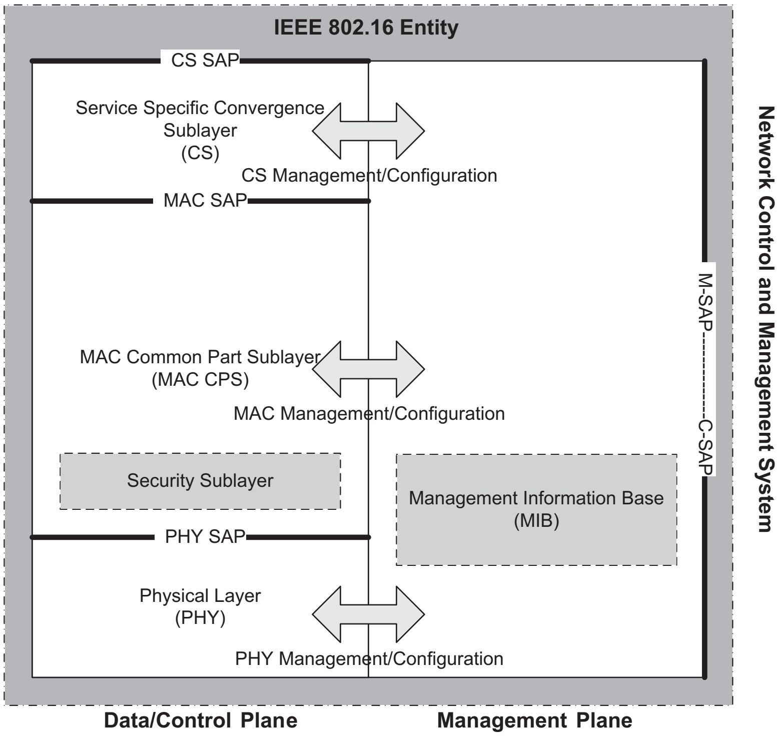

IEEE 802.16 Entity
CS SAP
Service Specific Convergence Sublayer (CS)
CS Management/Configuration
MAC SAP
MAC Common Part Sublayer (MAC CPS)
MAC Management/Configuration
Security Sublayer
Management Information Base (MIB)
PHY SAP
Physical Layer (PHY)
PHY Management/Configuration
M-SAP
C-SAP
Network Control and Management System
Data/Control Plane
Management Plane

FIGURE 3-3

The IEEE 802.16 reference model [3]

transformation or mapping of network-layer data packets into MAC SDUs. On the transmitter side, the CS receives the data packets through the CS Service Access Point (SAP) and delivers MAC SDUs to the MAC Common Part Sub-layer (MAC CPS) through the MAC SAP. This includes classifying network-layer SDUs and associating them with the proper MAC Service Flow Identifiers (SFID) and Connection Identifiers (CID). The convergence sub-layer also includes payload header suppression function to compress the higher-layer protocol headers. Multiple CS specifications are provided for interfacing with various network-layer protocols such as Asynchronous Transfer Mode (ATM)[i] and packet-switched protocols such as IP or Ethernet. The internal format of the CS payload is unique to the CS, and the MAC CPS is not required to understand the format of or parse any information from the CS payload.

The MAC CPS provides the core MAC functionality of system access, bandwidth allocation, connection establishment, and connection maintenance. It can receive data from the various convergence sub-layers, through the MAC SAP classified into particular MAC connections. An example of MAC CPS service definition is given in reference [3]. The Quality of Service (QoS) is further applied to the transmission and scheduling of data over the physical layer.

The MAC also contains a separate security sub-layer providing authentication, secure key exchange, and encryption. The user data, physical layer control, and statistics are transferred between the MAC CPS and the Physical Layer (PHY) via the PHY SAP which is implementation-specific.

The IEEE 802.16 physical layer protocols include multiple specifications, defined through several amendments and revisions, each appropriate for a particular frequency range and application. The IEEE 802.16 compliant devices include mobile stations or base stations. Given that the IEEE 802.16 devices may be part of a larger network, and therefore would require interfacing with entities for management and control purposes, a Network Control and Management System (NCMS) abstraction has been introduced in the IEEE 802.16 standard as a “black box” containing these entities [3]. The NCMS abstraction allows the physical and MAC layers specified in the IEEE 802.16 standard to be independent of the network architecture, the transport network, and the protocols used in the backhaul, and therefore would allow greater flexibility. The NCMS entity logically exists at both BS and MS sides of the radio interface. Any necessary inter-BS coordination is coordinated through the NCMS entity at the BS. An IEEE 802.16 entity is defined as a logical entity in an MS or BS that comprises the physical and MAC layers on the data, control, and management planes.

The IEEE 802.16f amendment (currently part of IEEE 802.16-2009 standard [3]) provided enhancements to IEEE 802.16-2004 standard, defining a management information base (MIB), for the physical and medium access control layers and the associated management procedures. The management information base originates from the Open Systems Interconnection Network Management Model and is a type of hierarchical database used to manage the devices in a communication network [5,6]. It comprises a collection of objects in a virtual database used to manage entities such as routers and switches in a network.

[i] Asynchronous Transfer Mode (ATM) is a packet switching protocol that encodes data into small fixed-sized cells and provides data link layer services that run over OSI layer 1, differing from other technologies based on packet-switched networks such as IP or Ethernet, in which variable-sized packets are used. ATM exploits properties of both circuit-switched and small packet-switched networks, making it suitable for wide area data networking, as well as real-time media transport. ATM uses a connection-oriented model and establishes a virtual circuit between two end-points before the actual data exchange begins [4].

The IEEE 802.16 standard describes the use of a Simple Network Management Protocol (SNMP),[ii] i.e., an IETF protocol suite, as the network management reference model. The standard consists of a Network Management System (NMS), managed nodes, and a service flow database. The BS and MS managed nodes collect and store the managed objects in the form of WirelessMAN Interface MIB and Device MIB that are made available to network management system via management protocols, such as SNMP. A Network Control System contains the service flow and the associated Quality of Service information that have to be provided to BS when an MS enters into the network. The Control SAP (C-SAP) and Management SAP (M-SAP) interface the control and management plane functions with the upper layers. The NCMS entity presents within each MS. The NCMS is a layer-independent entity that may be viewed as a management entity or control entity. Generic system management entities can perform functions through NCMS and standard management protocols can be implemented in the NCMS. If the secondary management connection does not exist, the SNMP messages, or other management protocol messages, may go through another interface in the customer premise or on a transport connection over the air interface. Figure 3-4 describes a simplified network reference

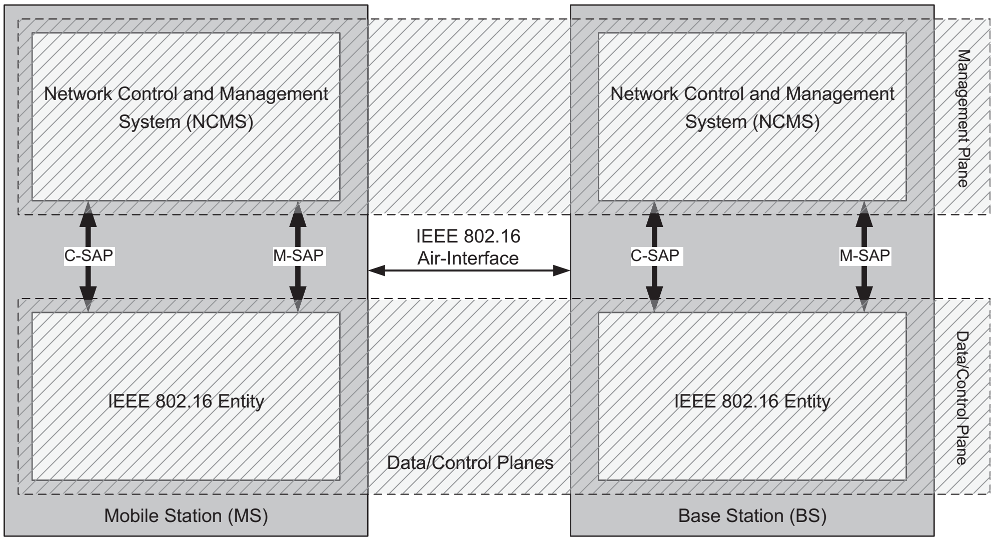


FIGURE 3-4

The IEEE 802.16 generic network reference model [3]

[ii] An SNMP-managed network consists of three key components: (1) a managed device; (2) an agent; and (3) a network management system. A managed device is a network node that contains an SNMP agent and resides in a managed network. Managed devices collect and store management information and make this information available to NMSs using SNMP. Managed devices, sometimes called network elements, can be any type of device including, but not limited to, routers, access servers, switches, etc. An agent is a network-management software module that resides in a managed device. An agent has local knowledge of management information and translates that information into a form compatible with SNMP. A network management system executes applications that monitor and control managed devices. The NMSs provide the processing and memory resources required for network management. One or more NMSs may exist on any managed network [7,8].

model. Multiple mobile stations may be attached to a BS. The MS communicates to the BS over the air interface using a primary management connection, basic connection or a secondary management connection [3]. The latter connection types have been replaced with new connection types in IEEE 802.16m standard [12].

3.1.1 The MS and BS Interface

The MAC management PDUs that are exchanged over the primary management connection[iii] can trigger, or are triggered by, primitives that are exchanged over either the C-SAP or the M-SAP, depending on the particular management or control operation. The messages that are exchanged over the secondary management connection can trigger or are triggered by primitives that are exchanged over the M-SAP. This interface is a set of SAP between an IEEE 802.16 entity and NCMS as shown in Figure 3-5. It consists of two parts: the M-SAP is used for delay-tolerant management-plane primitives; and the C-SAP is used for delay-sensitive control-plane primitives that support handovers,

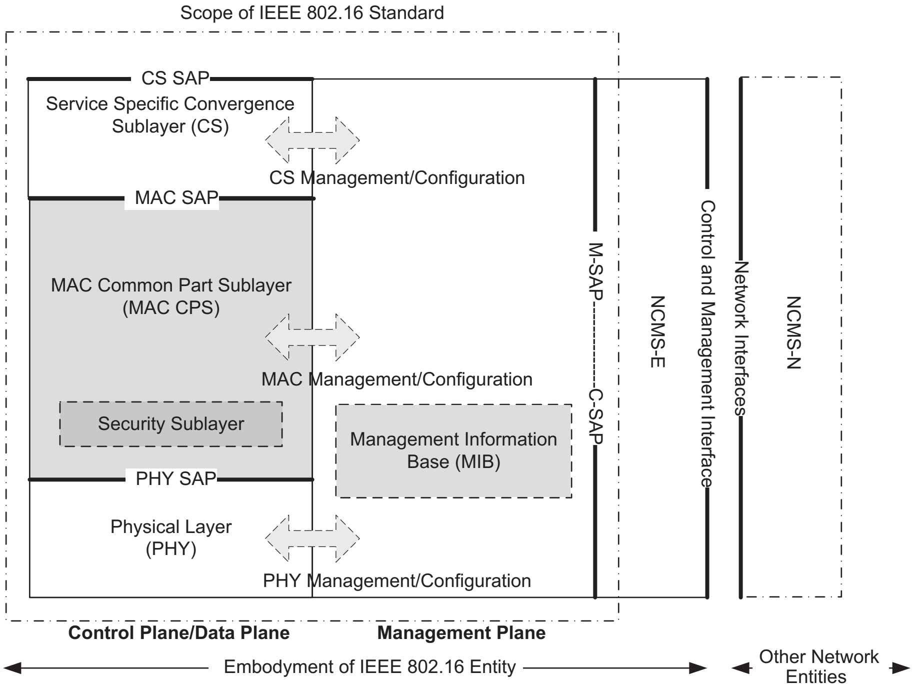

Scope of IEEE 802.16 Standard
CS SAP
Service Specific Convergence Sublayer (CS)
CS Management/Configuration
MAC SAP
MAC Common Part Sublayer (MAC CPS)
MAC Management/Configuration
Security Sublayer
Management Information Base (MIB)
PHY SAP
Physical Layer (PHY)
PHY Management/Configuration
M-SAP
C-SAP
NCMS-E
Control and Management Interface
Network Interfaces
NCMS-N
Control Plane/Data Plane
Management Plane
Embodiment of IEEE 802.16 Entity
Other Network Entities

FIGURE 3-5

Partitioning of the IEEE 802.16 network control and management system [3]

[iii] A management connection is used for transporting MAC management messages or standards-based messages. The primary management connection is established during network entry and is used to transport delay-tolerant MAC management messages. The secondary management connection may be established during MS registration that is used to transport standards-based messages; e.g., SNMP, DHCP messages.

security context management, radio resource management, and low power operations such as idle mode and paging functions.

3.1.2 Network Control and Management System

The Network Control and Management System is not part of the IEEE 802.16 standards, and is treated as a “black box.” It may be distributed with components residing on different nodes in a network. Part of the NCMS may be physically collocated with the IEEE 802.16 entity referred to as NCMS-E. The remaining part of the NCMS may be physically distributed across one or more network entities. This part of the NCMS is referred to as NCMS-N. Figure 3-5 shows the partitioning of the NCMS into NCMS-E and NCMS-N. The NCMS-E may have its own software platform and network protocol implementation, allowing it to communicate with external entities in the NCMS-N. The NCMS-E may provide an SNMP Agent compliant to IETF RFC3418 [13] and the SNMP/TCP/IP protocol stack, to allow for interactions with an SNMP manager. The NCMS-E may provide an Object Request Broker and implement a protocol stack to interact with components on other network entities within NCMS-N based on the CORBA architecture.[iv] The messages available to a manager in the NCMS-N are specified using Interface Description Language (IDL).[v] These messages encapsulate the interactions with the MIB. The IEEE 802.16 entity can be managed through Web Services.[vi] [11]

The decomposition of Network Control and Management System is depicted in Figure 3-6. These entities may be centrally located or distributed across the network. The exact functionality of these entities and their services is outside the scope of the IEEE 802.16 standard, but is shown here for illustration purposes and to allow description of the management and control procedures. The NCMS service manifestations on the MS and BS may have different configurations and functions.

The IEEE 802.16m reference model is very similar to that of the IEEE 802.16-2009 standard, with the exception of soft classification of MAC common part sub-layer into radio resource control and management and medium access control functions. As shown in Figure 3-7, this functional partitioning is logical, i.e., no SAP is required between the two classes of functions and no additional subheaders are appended to the SDUs. Furthermore, the functional elements on the data and control paths are explicitly classified into data- and control-plane functions. While similar functionalities exist in the IEEE 802.16-2009 standard, the functions and protocols are not explicitly categorized in the legacy standard except explicit separation of PHY, MAC CPS, and CS functions in the specification [3].

The categorization of the functions based on functional characteristics and relative position in the data/signaling processing path would ease analogy, and contrast with other radio access technologies

[iv] The Common Object Requesting Broker Architecture (CORBA) is a standard defined by the Object Management Group that enables software components written in multiple computer languages and running on multiple computers to work together [9].

[v] An Interface Description Language (IDL) is a specification language used to describe a software component’s interface. IDLs describe an interface in a language-independent way, enabling communication between software components that do not share a language; e.g., between components written in C++ and Java [10].

[vi] A Web Service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format. Other systems interact with the web service in a manner prescribed by its description using SOAP-messages (Simple Object Access Protocol is a lightweight protocol intended for exchanging structured information in a decentralized, distributed environment) typically conveyed using HTTP with an XML serialization in conjunction with other web-related standards.

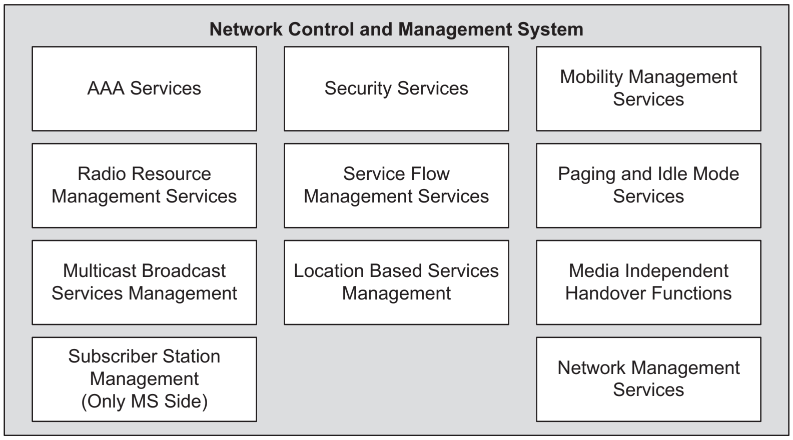

Network Control and Management System
AAA Services
Security Services
Mobility Management Services
Radio Resource Management Services
Service Flow Management Services
Paging and Idle Mode Services
Multicast Broadcast Services Management
Location Based Services Management
Media Independent Handover Functions
Subscriber Station Management (Only MS Side)
Network Management Services

FIGURE 3-6

Decomposition of the network control and management system [3]

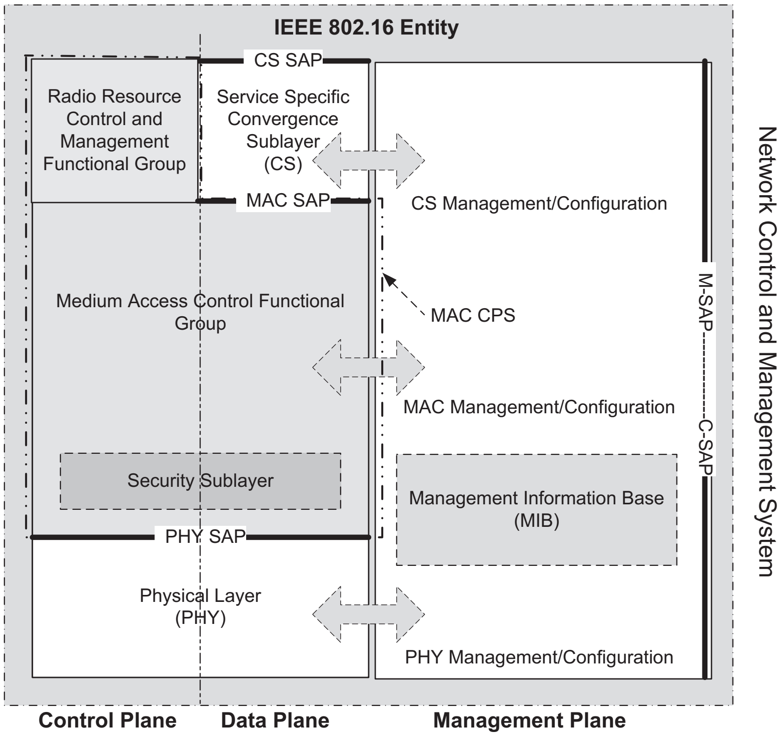

IEEE 802.16 Entity
CS SAP
Radio Resource Control and Management Functional Group
Service Specific Convergence Sublayer (CS)
MAC SAP
CS Management/Configuration
Medium Access Control Functional Group
MAC CPS
MAC Management/Configuration
Security Sublayer
Management Information Base (MIB)
PHY SAP
Physical Layer (PHY)
PHY Management/Configuration
M-SAP
C-SAP
Network Control and Management System
Control Plane
Data Plane
Management Plane

FIGURE 3-7

The IEEE 802.16m reference model [12]

such as 3GPP LTE/LTE-Advanced that have been designed based on similar structured protocol design methodology. Furthermore, the structured functional/protocol design in IEEE 802.16m would eliminate the inherent complexity and ambiguity of studying, understanding, and implementing the legacy standard.

It must be noted that there are new, modified or extended functions and protocols that are classified under generic classes of PHY, MAC CPS, and CS in IEEE 802.16m, where there are no counterparts in the legacy standard. Therefore, similarity of the reference models should not be interpreted as functional compatibility at the service access points. The backward compatibility of the IEEE 802.16m with the legacy standard ensures that non-compatible functions/protocols are not utilized in the time intervals where legacy base stations and mobile stations are supported in the network.

3.1.3 Data-Plane

The MAC and PHY functions of the IEEE 802.16m can be classified into three categories namely data-plane, control-plane, and management-plane. The data-plane (alternatively known as user-plane) comprises functions in the user data processing path, such as service flow classification and header compression, as well as MAC and PHY data packet processing and encryption functions. As shown in Figure 3-8, the IEEE 802.16m data-plane entity comprises the service specific Convergence Sub-layer,

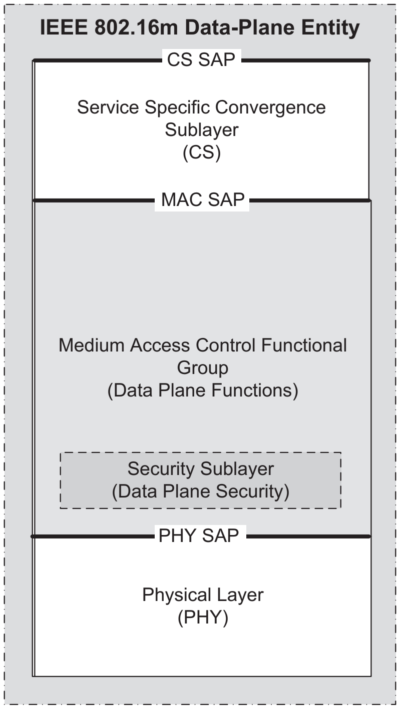

IEEE 802.16m Data-Plane Entity
CS SAP
Service Specific Convergence Sublayer (CS)
MAC SAP
Medium Access Control Functional Group (Data Plane Functions)
Security Sublayer (Data Plane Security)
PHY SAP
Physical Layer (PHY)

FIGURE 3-8

The IEEE 802.16m data-plane entity [12]

MAC functional group, security, and physical layer protocols corresponding to data-plane user packet processing. The MAC and PHY SAPs, while conceptually the same as those specified by IEEE 802.16-2009 standard, have different manifestation due to the new and/or modified MAC and PHY features introduced in the IEEE 802.16m standard. There is no significant change in CS SAP relative to that specified in the IEEE 802.16-2009 standard [3].

The CS provides a mapping of external network data formats (e.g., IP layer packets, ATM cells) received through CS SAP into MAC SDUs that are delivered to the MAC CPS via MAC SAP. This includes classifying external network SDUs and associating them with a proper Service Flow Identifier and Connection Identifier. A Service Flow (SF) is a MAC transport service that provides unidirectional transport of packets in the downlink or uplink [14]. It is identified by a 32-bit SFID. A service flow is characterized by a set of QoS parameters, i.e., a parameter set associated with a service flow identifier containing traffic parameters which define scheduling behavior of uplink or downlink service flows associated with transport connections. An admitted and active service flow is uniquely mapped to a CID. Note that the IEEE 802.16 standard supports two phase activation model. i.e., the resources for a service are first admitted or reserved and once the BS and MS negotiations are completed, the resources are activated [3].

The Generic Packet CS (GPCS) is a network layer protocol-agnostic packet convergence sub-layer that supports multiple network protocols over IEEE 802.16 air interface. The GPCS provides a generic packet convergence sub-layer. This layer uses the MAC SAP and exposes a SAP to GPCS applications. The GPCS does not redefine or replace other convergence sub-layers. Instead, it provides a SAP that is not protocol specific. With GPCS, packet parsing occurs above GPCS. The results of packet parsing are classification parameters provided to the GPCS SAP for parameterized classification; however, upper layer packet parsing is left to the GPCS application. With GPCS, the upper layer protocol that is immediately above the IEEE 802.16 GPCS is identified by a parameter known as GPCS protocol type. The GPCS protocol type is included in service flow management primitives and connection establishment messages. The GPCS defines a set of SAP parameters as the result of upper layer packet parsing. These are passed from upper layer to the GPCS in addition to the data packet. The SAP parameters include SFID, the MS MAC Address, data, and length. The GPCS allows multiplexing of multiple layer protocol types (e.g., IPv4, IPv6, and Ethernet) over the same IEEE 802.16 MAC connection. It is outside the scope of the GPCS protocol to specify how the upper layer multiplexes and de-multiplexes multiple protocol data packets over an IEEE 802.16 connection or service flow [3].

In multimedia streaming applications, the overhead of Internet Protocol (IP) [15], User Datagram Protocol (UDP) [16], and Real-time Transport Protocol (RTP) [17,18] payload headers are 40 bytes for IPv4 (or 60 bytes for IPv6 [19]). For voice-over-IP, this corresponds to approximately 60% of the total amount of encoded voice data (e.g., the RTP payload of 3GPP Adaptive Multi-Rate 12.2 kbps full-rate codec consists of 33 bytes [20]). Such large overheads may be tolerable in wired links where capacity is often not an issue, but are excessive for wireless systems where bandwidth is scarce. The IEEE 802.16 standard defines a native header compression algorithm that is part of the convergence sub-layer. The Payload Header Suppression (PHS) defined in the IEEE 802.16 standard compresses the repetitive or redundant parts of the payload header received from network layer. The PHS operation is based on the PHS rules, which provide all the parameters corresponding to header suppression of the SDU. Other standard header compression algorithms, such as Robust Header Compression (RoHC) defined by IETF [21], are also supported.

The RoHC scheme may be used as an alternative to PHS to compress the RTP/UDP/IP header of an IP packet. When RoHC is enabled for a service flow, the service flow constitutes what in IETF RFC 3095 is referred to as a RoHC channel [21,3]. Two service flows cannot share an RoHC channel, and two RoHC channels cannot share the same service flow. On a service flow for which RoHC has been enabled, all of the IP packet passes through the RoHC compressor on the transmitter side and the decompressor on the receiver side. The support of RoHC is negotiated between the BS and MS during capability negotiation [3].

The data-plane part of MAC CPS includes functions such as Automatic Repeat reQuest (ARQ), Packet Fragmentation/Packing, MAC PDU formation and encryption. The scheduler on the BS side allocates radio resources and multiplexes the users, and selects the appropriate MIMO mode, modulation and coding scheme based on the measurement reports that are received from the mobile stations. The ARQ is an error control mechanism at data link layer where the receiver may request the transmitter to resend a block of data that was erroneously detected or not received. An ARQ block is a distinct unit of data that is carried on an ARQ-enabled connection. Such a data unit is assigned a sequence number and is managed as a distinct entity by the ARQ state machines. The ARQ block size is a parameter that is negotiated during connection establishment. The ARQ mechanism may be disabled for some delay sensitive applications such as VoIP. Fragmentation is a process in which a MAC SDU is divided into one or more MAC SDU fragments. Packing is a process where multiple MAC SDUs are packed into a single MAC PDU payload. Both processes may be initiated by either a BS for a downlink connection or an MS for an uplink connection. Several MAC PDUs may be concatenated into a single transmission in the downlink or uplink.

The MAC PDUs containing user data are processed by the physical layer for over-the-air transmission. It must be noted that the physical layer processing of the user traffic can be different in terms of the permissible MIMO modes or modulation and coding schemes that are used.

3.1.4 Control-Plane

A set of Layer 2 control functions are needed to support various radio resource configuration, coordination, signaling, and management. This set of functions is collectively referred to as control-plane functions. The IEEE 802.16m control-plane entity comprises Radio Resource Control and Management (RRCM), MAC functional group, and physical layer protocols corresponding to control path. The RRCM functional class includes all control and management functions such as network entry/re-entry management, paging and idle mode management, multicast and broadcast service, etc. This group of functions is also known as Radio Resource Control (RRC) in other air interface standards such as 3GPP LTE. The MAC functional group consists of functions that perform physical layer control and signaling, scheduling services, QoS, etc. This functional group corresponds to Radio Link Control (RLC) and MAC layers in other air interface standards such as 3GPP LTE.

Figure 3-9 illustrates the IEEE 802.16m control-plane entity. As shown in this figure, the RRCM performs control and management of lower-layer functions. The control information is communicated with the mobile station via MAC management messages. The underlying functional elements of the RRCM sub-layer will be described in Section 3.2.

The security sub-layer in Figure 3-9 is shown with dotted line, since the IEEE 802.16m selectively encrypts and protects unicast MAC management messages. If the selective

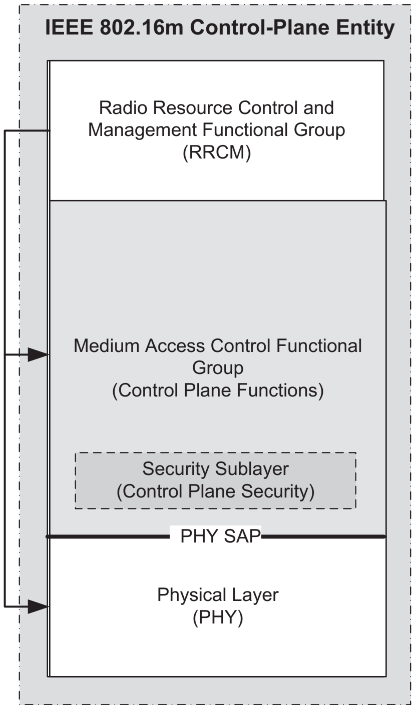

IEEE 802.16m Control-Plane Entity
Radio Resource Control and Management Functional Group (RRCM)
Medium Access Control Functional Group (Control Plane Functions)
Security Sublayer (Control Plane Security)
PHY SAP
Physical Layer (PHY)

FIGURE 3-9

The IEEE 802.16m control-plane entity [12]

confidentiality protection is utilized, the negotiated keying materials and cipher suites are used to encrypt the management messages. There are three levels of selective confidentiality protection applied to MAC management messages in the IEEE 802.16m: (1) no protection where the MS and BS have no shared security context or protection is not required, then the management messages are neither encrypted nor authenticated. Management messages before the authorization phase also fall into this category; (2) cipher-based message authentication code[vii] (CMAC) integrity protection protects the integrity of the entire MAC management message; and (3) advanced encryption standard-based[viii] (AES-CCM) authentication/encryption protects the integrity of payload and MAC header [3].

[vii] A cipher-based MAC (CMAC) is a block cipher-based message/code authentication algorithm. It may be used to provide assurance of the authenticity and integrity of user payloads. AES-CMAC provides stronger assurance of data integrity than a checksum or an error-detecting code. The verification of a checksum or an error-detecting code detects only accidental modifications of the data, while CMAC is designed to detect intentional, unauthorized modifications of the data, as well as accidental modifications.

[viii] The counter with CBC-MAC (CCM), as defined in IETF RFC 3610, is a generic authenticated encryption block cipher mode. CCM is only defined for use with 128-bit block ciphers, such as AES. It is an authenticated encryption algorithm designed to provide both authentication and privacy.

3.1.5 Management-Plane

A management-plane is also defined for external management and system configuration. Therefore, all management and protocol configuration entities, as well as management information base, fall into the management-plane category. Definition of management information bases are out of scope of the IEEE 802.16m standard. As shown in Figure 3-10, the management entity and the management information bases contained in the management-plane, configure and manage the functional entities in the data- and control-plane protocol layers. The IEEE 802.16 specification includes control and management SAPs as part of the management-plane that expose control-plane and management-plane functions to upper layers. Management-plane primitives and the C-SAP are used for more time sensitive control-plane primitives that support handovers, security context management, radio resource management, and low power system operations.

In addition, under the IEEE 802.16 standard, a user can be associated with a number of service flows, each characterized with different QoS parameters. This information is provisioned in a subscriber management system (e.g., AAA database) or a policy server. There are two service models: (1) static service model, where the subscriber station is not allowed to change the parameters of provisioned service flows or create new service flows dynamically; and (2) dynamic service model, where an MS or BS may create, modify or delete service flows dynamically. In the latter case,

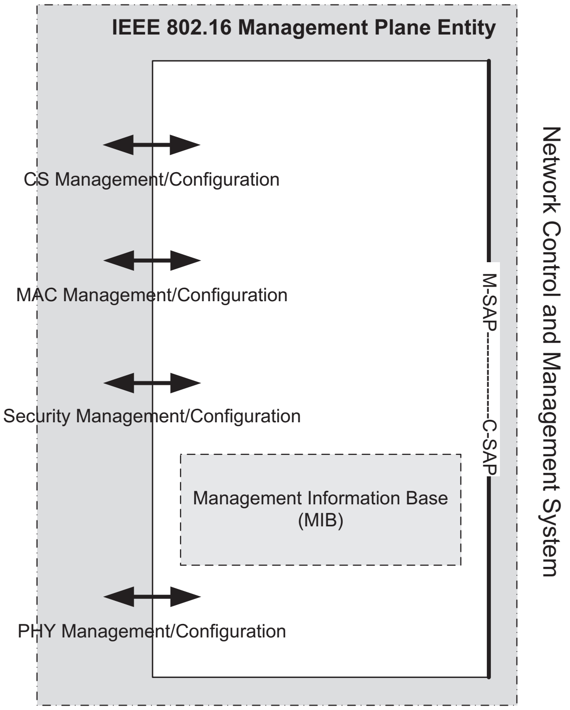

IEEE 802.16 Management Plane Entity
CS Management/Configuration
MAC Management/Configuration
Security Management/Configuration
Management Information Base (MIB)
PHY Management/Configuration
M-SAP
C-SAP
Network Control and Management System

FIGURE 3-10

The IEEE 802.16 management-plane entity [3] a dynamic service flow request is evaluated against the provisioned information to decide whether the request could be authorized. More precisely, the following steps are provisioned in the IEEE 802.16 specification for dynamic service flow creation [22]:

1. Permitted service flows and associated QoS parameters are provisioned for each subscriber via the management plane (management entity).
2. A service flow request initiated by the MS or BS is evaluated against the provisioned information, and the service flow is created if permissible.
3. A service flow thus created transitions to an admitted, and finally to an active, state due to BS action (this is possible under both static and dynamic service models). Transition to the admitted state involves the invocation of admission control in the BS and (soft) resource reservation, and transition to the active state involves actual resource assignment for the service flow. The service flow can directly transit from provisioned state to active state without going through admitted state.
4. A service flow can also transition in the reverse from an active to an admitted to a provisioned state.
5. A dynamically created service flow may also be modified or deleted.

As mentioned earlier, management information bases are collections of various network objects that are operated with the use of a Simple Network Management Protocol or SNMP [8,13]. The exact structure of the objects included in the management information base will depend on the configuration of the particular SNMP. However, additional extensions can allow for the addition of new objects outside the initial structure. Both the initial management information base and its extensions can be related to specific functions within a network. Some MIBs may be related to the definition of the domain name system, while other extensions may be associated with network objects like the fiber distributed data interface. While the initial management information base is usually defined as part of the SNMP, the extensions are generally set up as part of the basic management information base.

The Subscriber Station Management Primitives are a set of primitives to manage the status of mobile station. A management entity in the NCMS can change the status of mobile terminal. Those primitives are also used to notify the NCMS of information or events which are related to the status of the mobile terminal. The NCMS is a layer-independent entity that may be viewed as a management entity or control entity.

3.1.6 Service Access Point

A Service Access Point is defined as a reference point in a protocol stack where the services of a layer are available to its immediately neighboring layer. In other words, a SAP is a mapping between services of two neighboring layers. There are a number of SAPs in the IEEE 802.16 reference model (see Figure 3-7) that interface the adjacent protocol layers including PHY, MAC, and CS SAPs. The Management SAP may include primitives related to System configuration, Monitoring statistics, Notifications/Triggers, and Multi-mode interface management. The NCMS interacts with the MIB through the M-SAP. The Control SAP may include, but is not limited to, primitives related to handovers (e.g., notification of handover request from MS), idle mode mobility management (e.g., mobile station entering idle mode), subscriber and session management (e.g., mobile station requesting session set-up), radio resource management, AAA server signaling (e.g., *Extensible Authentication Protocol*[ix] payloads), Media Independent Handover (MIH)[x] services, and location detection and reporting capability. Unlike 3GPP LTE, the IEEE 802.16m does not explicitly define logical, transport, and physical channels (although the functionalities exist in the air interface protocols). In that case, the mapping between logical to transport and transport to physical channels would determine the SAPs between the corresponding layers.

3.1.7 Media-Independent Handover Reference Model for IEEE 802.16

The IEEE 802.21-2008 standard provides link-layer intelligence and other related network information to upper layers to optimize handovers between heterogeneous networks [23]. This includes media types specified by 3GPP, 3GPP2, and both wired and wireless media in the IEEE 802 family of standards. In the IEEE 802.21-2008 standard, media refers to the method or mode of accessing a telecommunication system (e.g., cable, radio, satellite), as opposed to sensory aspects of communication (e.g., audio, video). The standard addresses the support of handovers for both mobile and stationary users. For mobile users, handovers can occur when wireless link conditions change due to the users' movement. For the stationary user, handovers become imminent when the surrounding network environment changes, making one network more attractive than another. As an example, when making a network transition during a phone call, the handover procedures should be executed in such a way that any perceptible interruption to the conversation will be minimized. The standard supports cooperative use of information available at the mobile node and within the network infrastructure. The mobile node is well-positioned to detect available networks. The network infrastructure is well-suited to store overall network information, such as neighborhood cell lists, location of mobile nodes, and higher layer service availability. Both the mobile node and the network make decisions about connectivity. In general, both the mobile node and the network points of attachment (such as base stations and access points) can be multi-modal (i.e., capable of supporting multiple radio standards and simultaneously supporting connections on more than one radio interface).

Figure 3-11 shows the Media Independent Handover Function (MIHF), i.e., a function that realizes MIH services, for IEEE 802.16 based systems. The M-SAP and C-SAP are common between the MIHF and Network Control and Management System. The M-SAP specifies the interface between the MIHF and the management plane and allows MIHF payload to be encapsulated in management messages (such as MOB_MIH-MSG defined in the IEEE 802.16-2009 standard [3]). The primitives specified by M-SAP are used by a mobile node to transfer packets to a base station, both before and after it has completed the network entry procedures. The C-SAP specifies the interface between the MIHF and control-plane. M-SAP and C-SAP also transport MIH messages to peer MIHF entities. The CS-SAP is used to transfer packets from higher layer protocol entities after appropriate connections have been established with the network. The MIH-SAP specifies the interface of the MIHF with other higher layer entities such as transport layer, handover policy engine, and Layer 3 mobility protocols. In this model, C-SAP and M-SAP provide link services defined by MIH-LINK-SAP; C-SAP provides services before network entry; while CS-SAP provides services over the data-plane after network entry.

[ix] Extensible Authentication Protocol (EAP), as defined by IETF RFC 3748 and updated by IETF RFC 5247, is a universal authentication framework commonly used in wireless networks and point-to-point connections [24].

[x] Media Independent Handover (MIH) is a standard developed by the IEEE 802.21 to enable handover and interoperability between heterogeneous network types including both 802 and non-802 networks [23].

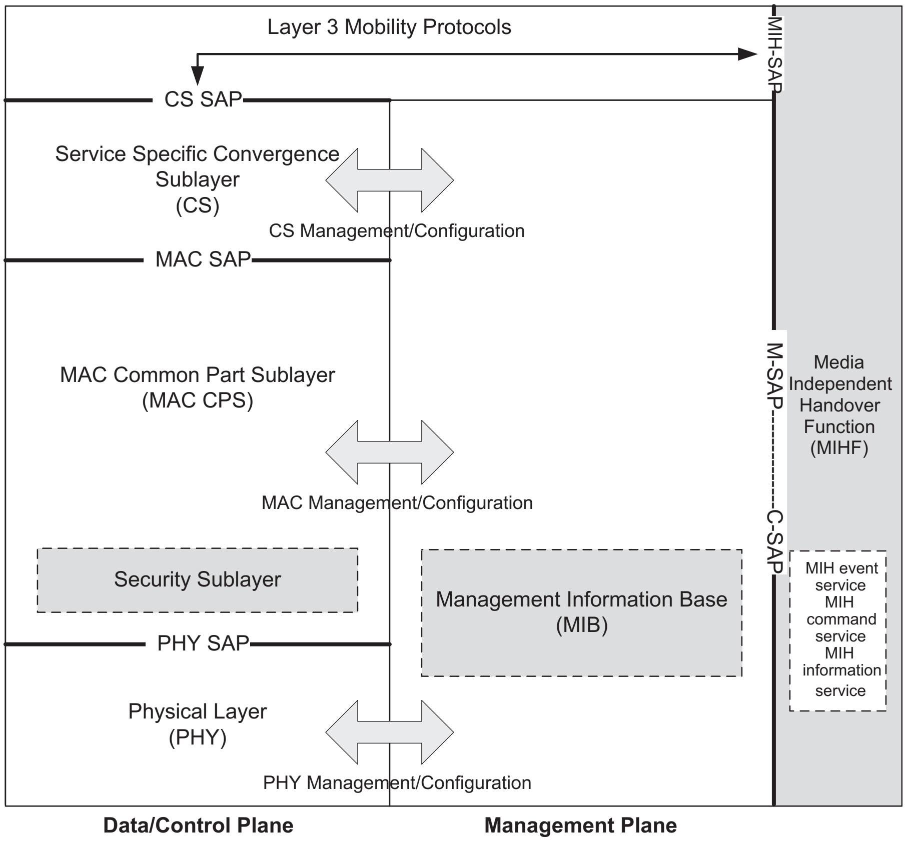

Layer 3 Mobility Protocols
MIH-SAP
CS SAP
Service Specific Convergence Sublayer (CS)
CS Management/Configuration
MAC SAP
MAC Common Part Sublayer (MAC CPS)
M-SAP
C-SAP
Media Independent Handover Function (MIHF)
MAC Management/Configuration
Security Sublayer
Management Information Base (MIB)
MIH event service
MIH command service
MIH information service
PHY SAP
Physical Layer (PHY)
PHY Management/Configuration
Data/Control Plane
Management Plane

FIGURE 3-11

MIH reference model for the IEEE 802.16 standard [23]

3.2 THE IEEE 802.16m PROTOCOL STRUCTURE

In this section, we further examine the functional elements of each protocol layer and their interactions. The 802.16m MAC common part sub-layer functions are classified into radio resource control and management functional group and medium access control functional group. The control-plane functions and data-plane functions are also separately classified. This would allow more organized, efficient, and structured method for specifying the MAC services in the IEEE 802.16m standard specification. As shown in Figure 3-12, the radio resource control and management functional group comprises several functional blocks including:

- Radio resource management block adjusts radio network parameters related to the traffic load, and also includes the functions of load control (load balancing), admission control, and interference control;

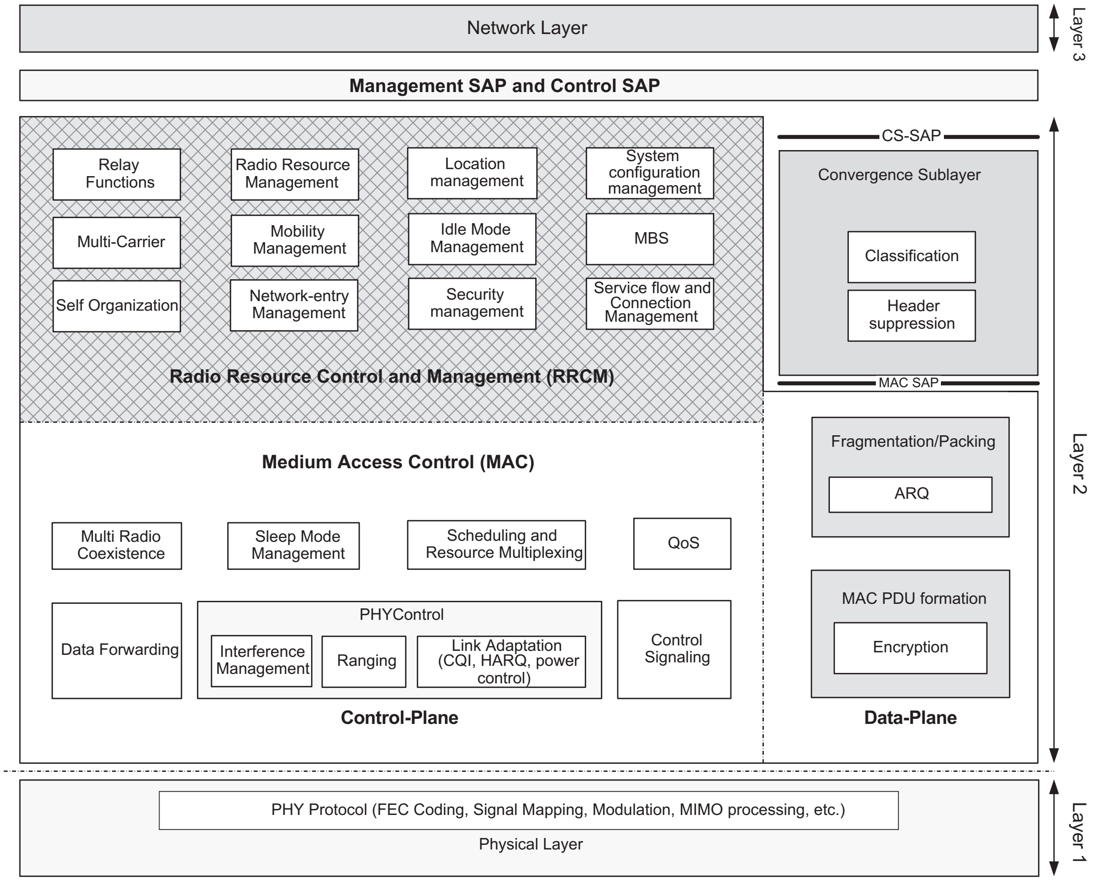


FIGURE 3-12

The IEEE 802.16m general protocol stack [12]

- Mobility management block scans neighbor BSs and decides whether MS should perform handover operation;
- Network-entry management block controls initialization and access procedures and generates management messages during initialization and access procedures;
- Location management block supports location based service (LBS), generates messages including the LBS information, and manages location update operation during idle mode;
- Idle mode management block controls idle mode operation, and generates the paging advertisement message based on paging message from paging controller in the core network;
- Security management block performs key management for secure communication. Using managed key, traffic encryption/decryption and authentication are performed;
- System configuration management block manages system configuration parameters, and generates broadcast control messages such as superframe headers;

- Multicast and broadcast service (MBS) block controls and generates management messages and data associated with MBS;
- Service flow and connection management block allocates Station Identifier (STID) and Flow Identifiers (FIDs) during access/handover service flow creation procedures.

The medium access control functional group, on the control plane, includes functional blocks which are related to physical layer and link controls such as:

- PHY control block performs PHY signaling such as ranging, channel quality measurement/feedback (CQI), and HARQ ACK or NACK signaling;
- Control signaling block generates resource allocation messages such as advanced medium access protocol, as well as specific control signaling messages;
- Sleep mode management block handles sleep mode operation and generates management messages related to sleep operation, and may communicate with the scheduler block in order to operate properly according to sleep period;
- Quality-of-service block performs rate control based on QoS input parameters from connection management function for each connection;
- Scheduling and resource multiplexing block schedules and multiplexes packets based on properties of connections.

The MAC functional group on the data-plane includes functional blocks such as:

- Fragmentation/packing block performs fragmentation or packing of MAC Service Data Units (MSDU) based on input from the scheduling and resource multiplexing block;
- Automatic Repeat Request block performs MAC ARQ function. For ARQ-enabled connections, a logical ARQ block is generated from fragmented or packed MSDUs of the same flow and sequentially numbered;
- MAC protocol data unit formation block constructs MAC PDU (MPDU) such that BS/MS can transmit user traffic or management messages into PHY channels.

The IEEE 802.16m protocol structure is similar to that of the IEEE 802.16, with some additional functional blocks in the control-plane for new features including the following:

- Relay functions enable relay functionalities and packet routing in relay networks.
- Self organization and self-optimization functions enable home BS or femto-cells and plug-and-play form of operation for indoor BS (i.e., femto-cell[xi]).
- Multi-carrier functions enable control and operation of a number of adjacent or non-adjacent RF carriers (i.e., virtual wideband operation) where the RF carriers can be assigned to unicast and/or multicast and broadcast services. A single MAC instantiation will be used to control several physical layers. The mobile terminal is not required to support multi-carrier operation. However, if it does support multi-carrier operation, it may receive control and signaling, broadcast, and synchronization channels through a primary carrier and traffic assignments (or services) via the secondary carriers.

[xi] Femto-cells are low-power wireless access points that operate in licensed spectrum to connect standard mobile devices to a mobile operator's network using residential DSL or cable broadband connections [25,26].

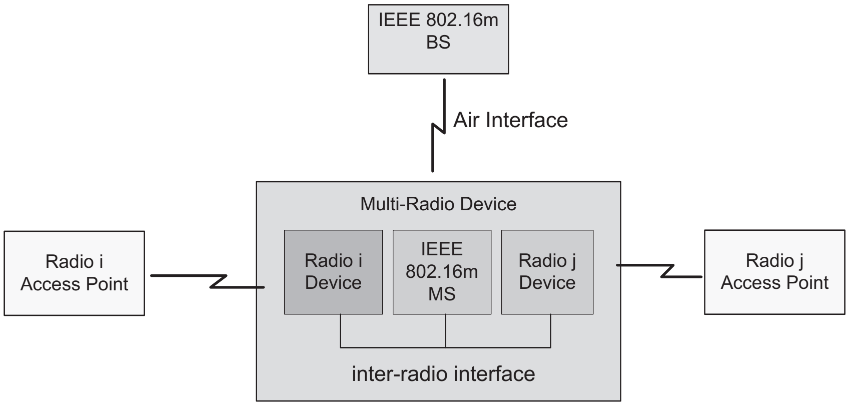


FIGURE 3-13

A generic multi-radio coexistence model [12]

- Multi-radio coexistence functions in IEEE 802.16m enable the MS to generate MAC management messages in order to report information on its collocated radio activities, and enable the BS to generate MAC management messages to respond with the appropriate actions to support multi-radio coexistence operation. Furthermore, the multi-radio coexistence functional block at the BS communicates with the scheduler functional block to assist proper scheduling of the MS according to the reported collocated coexistence activities. The multi-radio coexistence function is independent of the sleep mode operation to enable optimal power efficiency with a high level of coexistence support. However, when sleep mode provides sufficient collocated coexistence support, the multi-radio coexistence function may not be used (see Figure 3-13).
- Interference management functions are used to manage the inter-cell/sector interference effects. The procedures include MAC layer functions (e.g., interference measurement/assessment reports sent via MAC signaling and interference mitigation by scheduling and flexible frequency reuse), and PHY functions (e.g., transmit power control, interference randomization, interference cancellation, interference measurement, transmit beamforming/precoding). The inter-BS coordination functions coordinate the operation of multiple base stations by exchanging information, about interference statistics between the base stations via core-network signaling.

3.2.1 Data-Plane and Control-Plane Functions in Base Stations and Mobile Stations

Figure 3-14 shows the user data processing path at the BS and MS. As shown in the figure, the user data traverses the path from network layer to physical layer and *vice versa*. In the transmitter side, a network layer packet is processed by the convergence sub-layer, the ARQ function (if enabled), the fragmentation/packing function, and the MAC PDU formation function, to form the MAC PDU to be sent to the physical layer for processing. In the receiver side, a physical layer SDU is processed by MAC PDU formation function, the fragmentation/packing function,

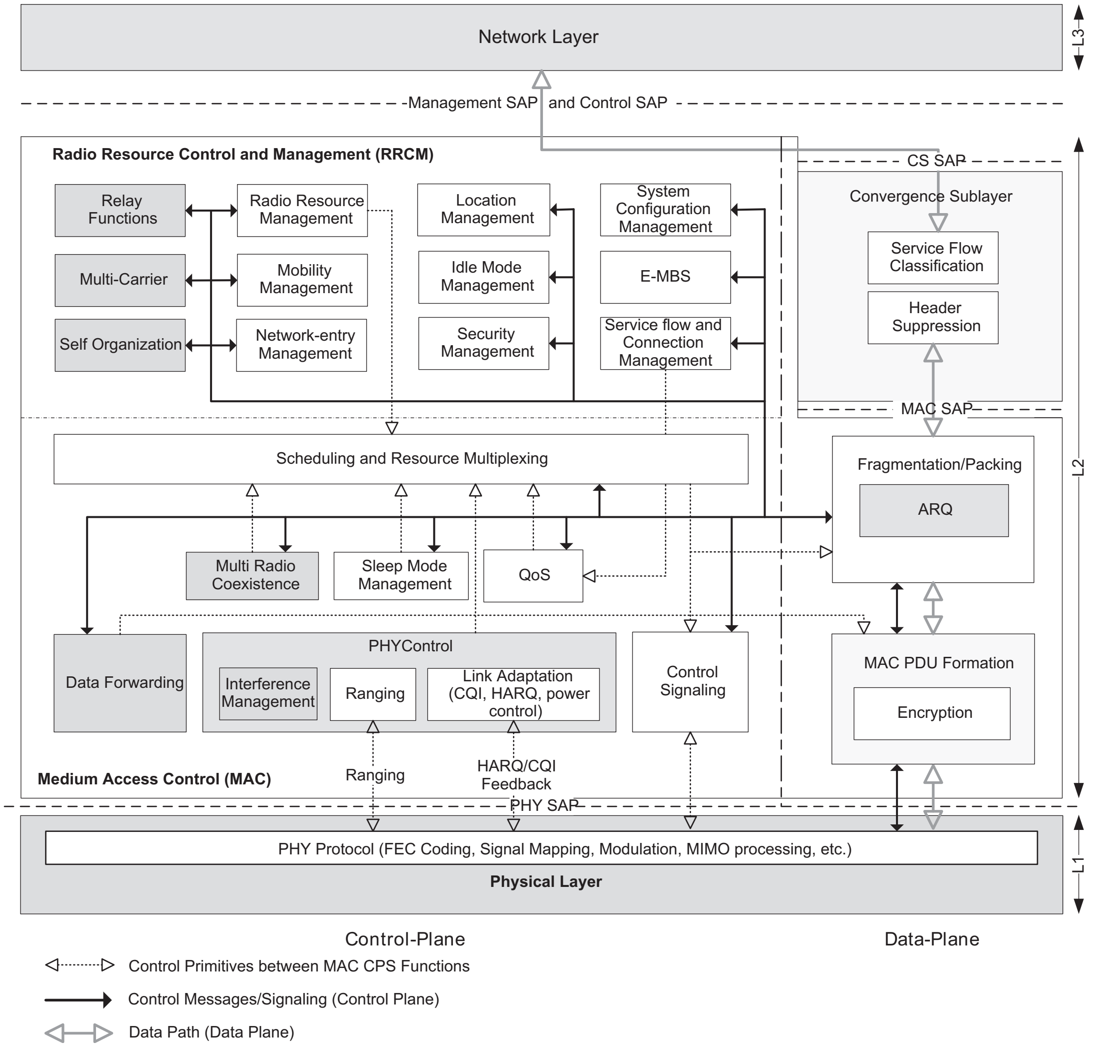


FIGURE 3-14

Signal flow graph in data- and control-planes [12]

the ARQ function (if enabled), and the convergence sub-layer function, to form the network layer packets. The control primitives between the MAC CPS functions and between the MAC CPS and PHY that are related to the processing of user traffic data are also shown in Figure 3-14.

The control-plane signaling and processing flow graph at the BS and the MS are shown in Figure 3-14. In the transmitter side, the flow of control primitives from control-plane functions to data-plane functions and processing of control-plane signals by data-plane functions in order to construct MAC management messages and MAC header/sub-headers, to be transmitted over the air interface, are illustrated. In the receiver side, the arrows show the processing of the MAC control messages through data-plane functions and the reception of the corresponding control-plane signals by control-plane functions. The dotted arrows show the control primitives between MAC CPS functions and between MAC CPS and physical layer functions that are related to the processing of control-plane signaling. The control primitives to/from M-SAP/C-SAP define the network related functionalities, such as inter-BS interference management, inter/intra RAT mobility management, etc., as well as management-related functionalities, such as location management, system configuration, etc.

Figure 3-15 illustrates the IP packet processing in an IEEE 802.16m base station transmitter and mobile station receiver. The main functional components of each layer and their interconnections are identified. It is further shown how the MAC scheduler in the base station, based on the periodic reports and measurements provided by each mobile station, generates appropriate control signals to select the best modulation and coding scheme, re-transmission method and number of re-transmissions, MIMO mode, and antenna configuration according to the channel conditions that the mobile station is experiencing.

3.2.2 Data-Plane and Control-Plane Functions in Relay Stations

Multi-hop relay is an optional entity that may be deployed in conjunction with base stations to provide additional coverage or performance improvements in a radio access network. In relay-enabled networks, the BS may be replaced by a multi-hop relay BS (i.e., a BS that supports relay capability over the relay links) and one or more relay stations (RS). The traffic and signaling between the mobile station and relay-enabled BS are relayed by the RS, thus extending the coverage and performance of the system in areas where the relay stations are deployed. Each RS is under the control of a relay-enabled BS [27].

In a multi-hop relay system, the traffic and signaling between an access RS and the BS may also be relayed through intermediate relay stations. The RS may either be fixed in location or it may be mobile. The mobile station may also communicate directly with the serving BS. The various relay-enabled BS features defined in the IEEE 802.16j-2009 standard allow a multi-hop relay system to be configured in several modes. The air interface protocols, including the mobility features on the access link (i.e., RS-MS link), remain unchanged.

The IEEE 802.16j-2009 standard specified a set of new functionalities on the relay link to support the RS–BS communication. Two different modes; i.e., centralized and distributed scheduling modes, were specified for controlling the allocation of bandwidths for an MS or an RS. In centralized scheduling mode the bandwidth allocation for subordinate mobile stations of an RS is determined at the serving BS. On the other hand, in distributed scheduling mode the bandwidth allocation of the subordinate stations is determined by the RS, in cooperation with the BS. Two different types of RS are defined, namely transparent and non-transparent. A non-transparent RS can operate in both centralized and distributed scheduling mode, while a transparent RS can only operate in centralized scheduling mode. A transparent RS communicates with the base station and subordinate mobile stations using the same carrier frequency. A non-transparent RS may communicate with the base station and the subordinate mobile stations via the same or different carrier frequencies.

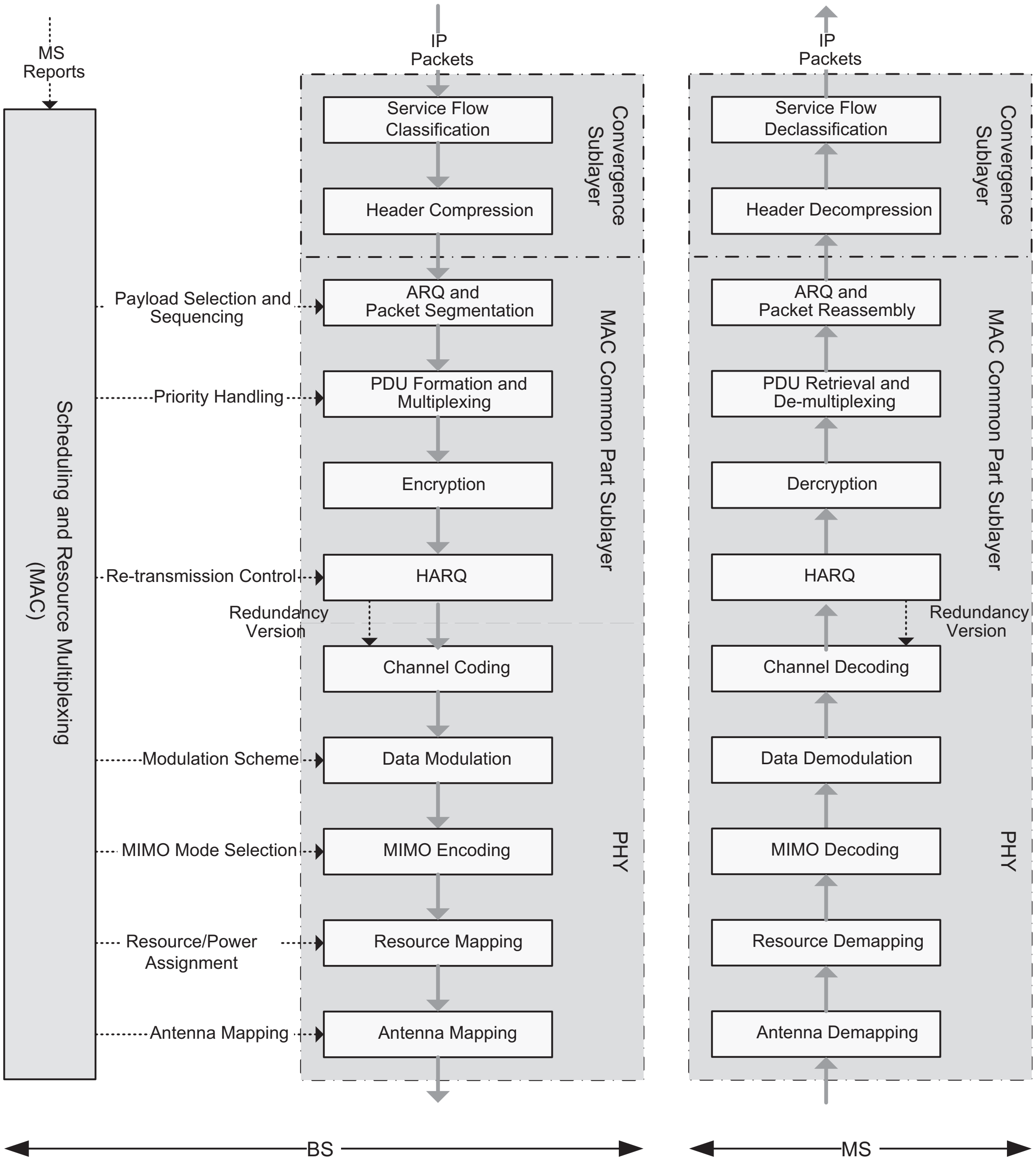


FIGURE 3-15

IP Packet processing and retrieval in the BS and MS

Relaying in the IEEE 802.16m system is performed using a decode-and-forward paradigm and supports TDD and FDD duplex modes. In TDD deployments, the relay stations operate in time-division transmit and receive (TTR) mode,[xii] whereby the access and relay link communications are multiplexed using time division multiplexing over a single RF carrier. In the IEEE 802.16m system, the relay stations operate in non-transparent mode, which essentially means that the relay stations compose and transmit the synchronization channels, system information, and the control channels for the subordinate stations. In any IEEE 802.16m deployment supporting relay functionality, a distributed scheduling model is used where each infrastructure station (BS or RS) schedules the radio resources on its subordinate links. In the case of a relay station, the scheduling of the resources is within the radio resources assigned by the BS. The BS notifies the relay and mobile stations of the frame structure configuration. The radio frame is divided into access and relay zones. In the access zone, the BS and the RS transmit to, or receive from, the mobile stations. In the relay zone, the BS transmits to the relay and the mobile stations, or receives from the relay and mobile stations. The start times of the frame structures of the BS and relay stations are aligned in time. The BS and relay stations transmit synchronization channels, system information, and the control channels to the mobile stations at the same time.

The MAC layer of a relay station includes signaling extensions to support functions such as network entry of an RS and of an MS through an RS, bandwidth request, forwarding of PDUs, connection management, and handover. Two different security modes are defined in the IEEE 802.16j-2009 standard: (1) a centralized security mode that is based on key management between the BS and an MS; and (2) a distributed security mode which incorporates authentication and key management between the BS and a non-transparent access RS, and between the access-RS and an MS. An RS may be configured to operate either in normal CID allocation mode, where the primary management, secondary, and basic CIDs are allocated by the BS, or in local CID allocation mode where the primary management and basic CID are allocated by the RS.

The IEEE 802.16m RS uses the same security architecture and procedures as an MS to establish privacy, authentication, and confidentiality between itself and the BS on the relay link. The IEEE 802.16m relay stations use a distributed security model. The security association is established between an MS and an RS during the key exchange similar to a macro BS. The RS uses a set of active keys shared with the MS to perform encryption/decryption and integrity protection on the access link. The RS runs a secure encapsulation protocol with the BS based on the primary security association. The access RS uses a set of active keys shared with the BS to perform encryption/decryption and integrity protection on the relay link. The MAC PDUs are encapsulated within one relay MAC PDU and are encrypted or decrypted by primary security association, which is established between the RS and the BS. The security contexts used for the relay link (between a BS and an RS) and the access links (between an RS and an MS) are different and are maintained independently. The key management is the same as that performed by a macro BS.

Figure 3-16 shows the IEEE 802.16m relay station protocol stack. An RS may consist of a subset of the protocol functions shown in Figure 3-16; however, the ingredients of each subset of functions depend on the type or category of the RS, as well as other deployment requirements. The functional

[xii] Time-division transmit and receive is a relay mechanism where transmission to subordinate station and reception from the super-ordinate station or transmission to the super-ordinate station and reception from the subordinate station is separated in time.

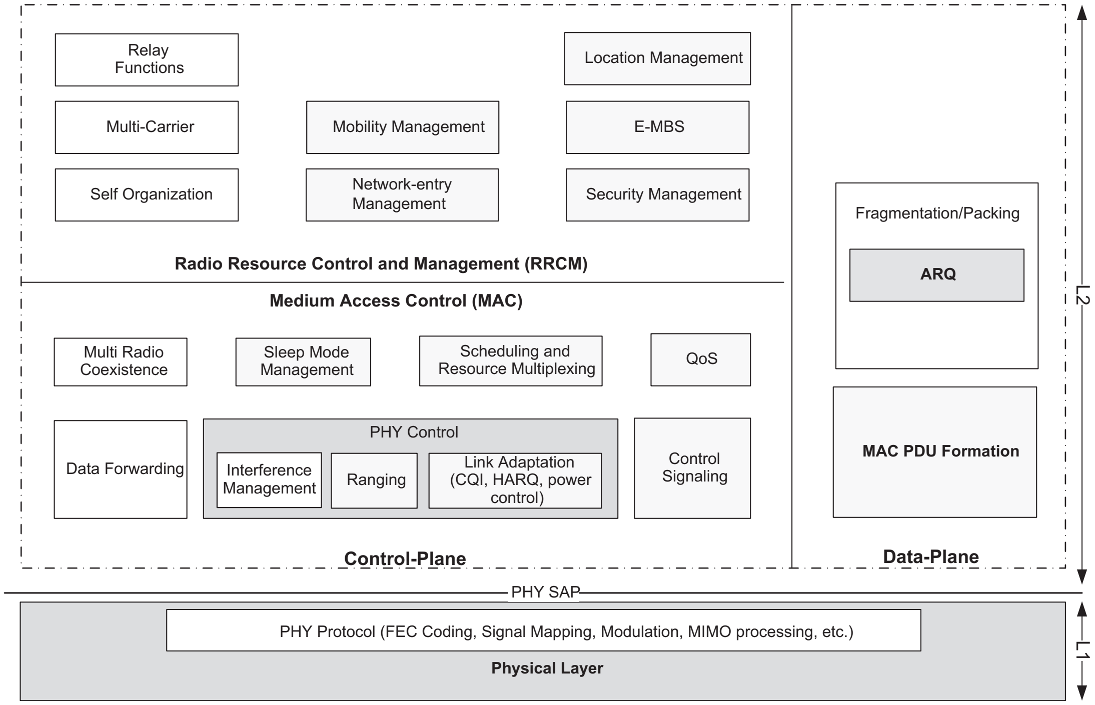


FIGURE 3-16

Protocol structure of the IEEE 802.16m relay stations [12]

blocks and the definitions provided in this section do not imply their support in all IEEE 802.16m RS implementations. The IEEE 802.16m relay capabilities are partially (and conceptually) based on the functionalities and features specified in the IEEE 802.16j-2009 standard [27]. The IEEE 802.16j-2009 standard does not define a system profile; therefore, many non-interoperable realizations of the relay stations can be considered. The WiMAX Forum technical working group started (and later abandoned) an initiative to define a relay system profile to facilitate certification and deployment of interoperable relay stations in mobile WiMAX systems. The non-transparent relay stations perform the same functions as a regular base station; however in some usage models, some functionalities of the regular base station may not be implemented in the relay station, resulting in less complexity and lower cost of implementation and deployment.

The IEEE 802.16m RS MAC CPS is divided into two sub-layers: (1) Radio Resource Control and Management sub-layer; and (2) Medium Access Control sub-layer. The RS RRCM sub-layer includes the following functional blocks that are related to the RS radio resource management functions:

- Mobility management;
- Network-entry management;
- Location management;
- Security management;

- Multicast and broadcast service;
- Relay functions;
- Self organization;
- Multi-carrier operation.

In Figure 3-16, the mobility management block supports the MS handover operation in cooperation with the BS. The network-entry management block performs RS/MS initialization procedures, as well as the RS network entry/attachment procedure to the BS. The network-entry management block may generate management messages needed during RS/MS initialization procedures and performing the network entry. The location management block supports location-based services including positioning data at the RS and reporting location information to the BS. The security management block performs the key management functions for the RS. Since an IEEE 802.16m relay uses a distributed security model, there are two sets of security protocols on the access and the relay links.

The enhanced multicast and broadcast service block is responsible for coordination, scheduling, and distribution of the E-MBS content to the subscribed users in the relay coverage area. The relay functional block includes procedures to maintain relay paths. The self-organization block performs functions to support the RS self-configuration and the RS self-optimization mechanisms which are coordinated by the BS. These functions include procedures to request the relay stations or mobile stations to report measurements for self-configuration and self-optimization, receive measurements from the relay stations or mobile stations, and report measurements to the BS. These functions also include procedures to adjust the RS parameters and configuration for self-configuration and/or optimization with or without coordination with the BS.

The multi-carrier operation block enables a common MAC entity to control a physical layer that may span over multiple frequency channels at the RS. The RS MAC sub-layer includes the following functional blocks which are related to the physical layer and link control:

- Physical layer control;
- Control signaling;
- Sleep mode management;
- Quality of service;
- Scheduling and resource multiplexing;
- ARQ function;
- Fragmentation/packing;
- MAC PDU formation;
- Data forwarding;
- Multi-radio coexistence.

As shown in Figure 3-16, the physical layer control block manages signaling schemes such as ranging, measurement, reporting, and HARQ feedback at the RS. Based on CQI and HARQ feedback, the physical layer control block estimates channel conditions of RS/MS and performs link adaptation. The control signaling block performs the RS resource allocation and generates control messages. The sleep mode management block manages sleep mode operation of mobile stations serviced by the RS in coordination with the BS. The QoS block performs rate control according to QoS parameters. The scheduling and resource multiplexing block which resides in the RS is used to support distributed scheduling, and schedules the transmission of MAC PDUs. The ARQ block assists MAC ARQ functions between the BS and RS over the relay link and between MS and RS over the access link. The MAC SDUs may be fragmented or augmented depending on the size of the payloads. The fragmentation/packing block in the RS side includes the unpacking and repacking of data fragments that have been received for relaying, in order to adapt the size of MAC PDUs to the estimated channel quality of the outbound link.

The MAC PDU formation block constructs MAC PDUs which contain user traffic or management messages. User traffic is assumed to have originated at either the BS or MS. The MAC PDU construction block may add or modify MAC PDU control information (e.g., MAC header). The data forwarding block performs routing functions on the link between the BS and the RS or MS. The data forwarding block may work in conjunction with other blocks such as scheduling and resource multiplexing and MAC PDU formation.

The interference management block at the RS performs inter-cell and inter-RS interference management. This function includes reception of interference level measurements and selection of transmission format used for the mobile stations attached to the RS. The control functions can be divided among the serving BS and the relay stations using a centralized model or a distributed model. In a centralized model, the serving BS generates control signals, and the relay stations communicate the control information between the BS and MS. In a distributed model, the RS generates control signals for the subordinate mobile stations and makes the BS aware of that control information. The determination of whether a particular control function should be centralized or distributed is made independently for each control function.

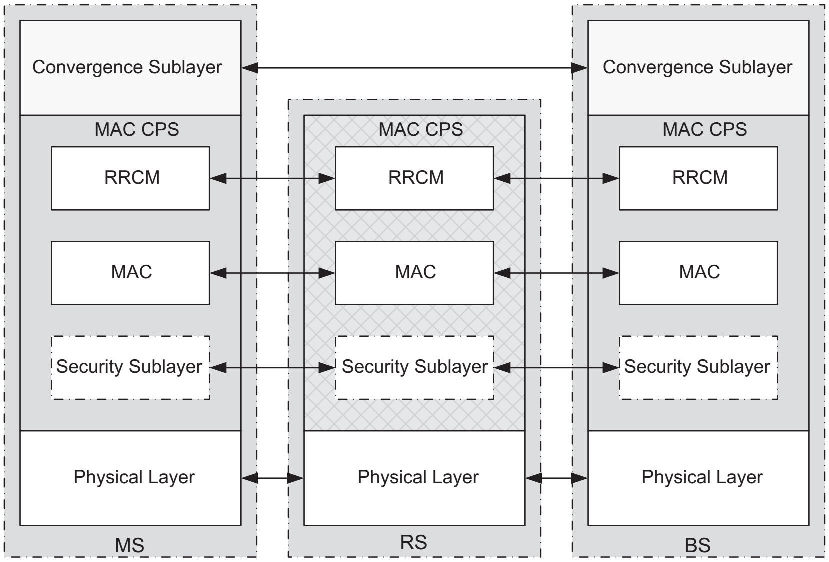


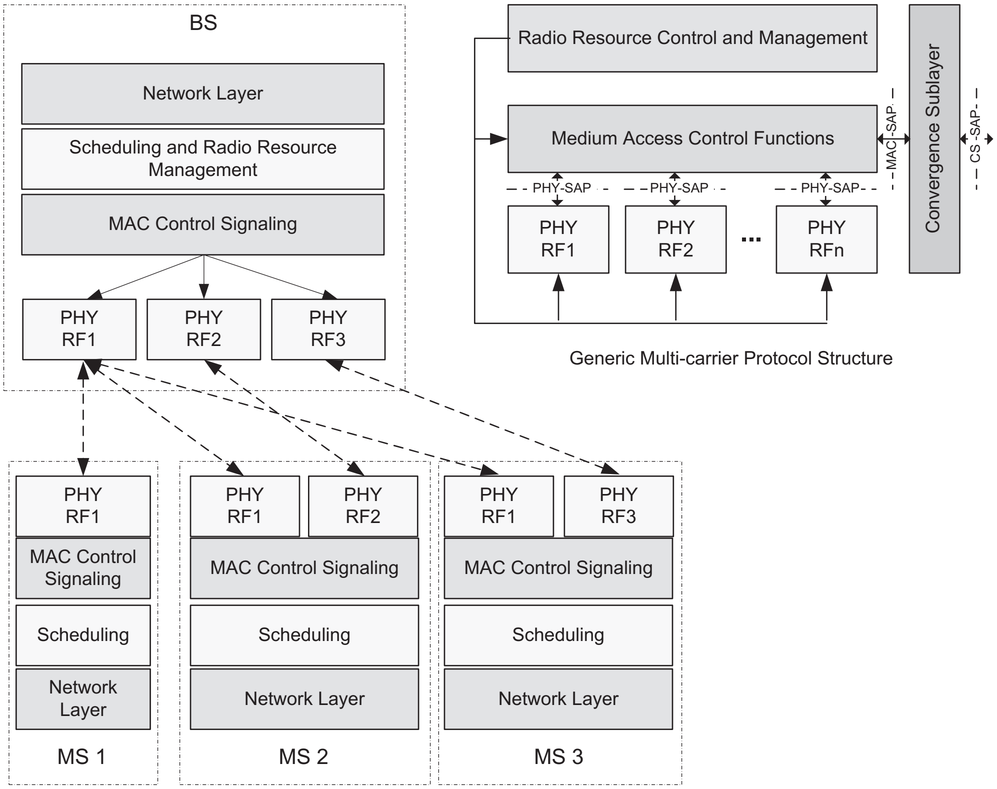
FIGURE 3-17

Protocol termination in a relay-enabled network

The multi-radio coexistence block within the RS coordinates collocated multi-radio operation of the subordinate mobile stations in coordination with the BS. Based on the earlier description of the functions and protocols performed by a relay station, the control- and data-plane protocol termination is shown in Figure 3-17. The convergence sub-layer protocols are terminated at the MS and the BS. However, some of the MAC CPS protocols are terminated at the RS. Due to the use of a distributed security model, the security functions including encryption and packet validation are terminated at the RS on the relay and access links.

3.2.3 Protocol Structure for Support of Multi-Carrier Operation

The generic protocol structure for support of multi-carrier operation is illustrated in Figure 3-18. A single MAC instance controls a number of physical layers spanning over multiple frequency bands. Some MAC messages transmitted over one RF carrier may also apply to other RF carriers. The RF channels may be of different bandwidths (e.g., 5, 10, and 20 MHz), and can be contiguous or non-



Generic Multi-carrier Protocol Structure

Example Multi-carrier Operation (Mobile Stations with Different Capabilities)

FIGURE 3-18

Multi-carrier operation using a single MAC instantiation [12] contiguous in frequency. The RF channels may support different duplexing schemes, e.g., frequency division duplex (FDD) mode, time division duplex (TDD) mode, or a combination of multicast and/or unicast RF carriers [12]. As shown in Figure 3-18, the MAC entity can provide simultaneous service to mobile stations with different bandwidth capabilities, such as operation over one RF channel at a time or aggregation across contiguous or non-contiguous frequency bands.

3.2.4 Protocol Structure for Support of Multicast and Broadcast Services

Multicast and broadcast service is a point-to-multipoint communication scheme where data packets are transmitted simultaneously from a single source to multiple destinations. The term broadcast refers to the ability to deliver content to all users. Multicast, on the other hand, refers to distribution of content among a specific group of users that are subscribed to those services. The multicast and broadcast content is transmitted over a geographical area referred to as a zone. An MBS zone is a collection of one or more base stations transmitting the same content. Each BS capable of MBS service may belong to one or more MBS zones. Each MBS zone is identified by a unique zone identifier [3].

An MS can receive the MBS content within the MBS zone in connected state or idle state. A BS may provide multicast and broadcast services corresponding to different MBS zones. The MBS data bursts may be transmitted in the form of several sub-packets, and these sub-packets may be transmitted in different time intervals to allow the MS to combine the sub-packets without transmission of acknowledgement. The mobile stations in an MBS zone are assigned a common multicast station identifier. The IEEE 802.16m supports two types of MBS access: (1) single-BS; and (2) multi-BS. The single-BS access is implemented over multicast and broadcast transport connections within one BS, while multi-BS access is realized by transmitting MBS data through multiple base stations. The MBS PDUs are transmitted by all base stations in the same MBS zone. That transmission is supported either in the non-macro diversity mode or macro diversity mode. An MBS zone may be formed by only one

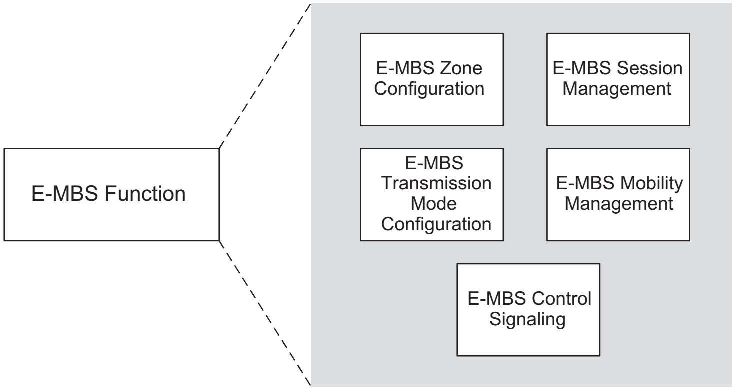


FIGURE 3-19

Breakdown of the E-MBS function (control-plane) [12]

BS. The MS may support both single-BS and multi-BS access. The MBS service may be delivered via a dedicated RF carrier or a mixed unicast, multicast, and broadcast RF carrier.

The IEEE 802.16m Enhanced Multicast and Broadcast Service (E-MBS) consists of MAC and PHY protocols that define interactions between the mobile stations and the base stations. While the basic definitions of IEEE 802.16m E-MBS are consistent with that of the IEEE 802.16-2009 standard [3], some enhancements and extensions are incorporated to provide improved functionality and performance [12]. The breakdown of E-MBS MAC function into constituent components is shown in Figure 3-19. In the control-plane, E-MBS MAC function operates in conjunction with other unicast service MAC functions. The unicast MAC functions may operate independently from E-MBS MAC function. The E-MBS MAC function may operate differently depending on whether it is operating in active mode or idle mode [12].

The E-MBS MAC function consists of the following sub-blocks:

- E-MBS Zone Configuration: this function manages the configuration and advertisement of E-MBS zones. A BS may belong to multiple E-MBS zones.
- E-MBS Transmission Mode Configuration: this function describes the transmission mode in which E-MBS is delivered over the air interface such as single-BS and multi-BS transmission.
- E-MBS Session Management: this function manages E-MBS service registration and deregistration and session start, update, or termination.
- E-MBS Mobility Management: this block manages the zone update procedures when an MS crosses the E-MBS zone boundary.
- E-MBS Control Signaling: this block broadcasts the E-MBS scheduling and physical channel mapping to facilitate E-MBS reception and power saving.

3.3 3GPP LTE/LTE-ADVANCED PROTOCOL STRUCTURE

In this section, we provide an overview of 3GPP LTE/LTE-Advanced protocol structure. Figures 3-20 and 3-21 illustrate the user-plane (U-plane) and control-plane (C-plane) protocol stacks, respectively. In the C-plane, the Non-Access Stratum (NAS) functional block is used for network attachment, authentication, setting up bearers, and mobility management. All NAS messages are ciphered and

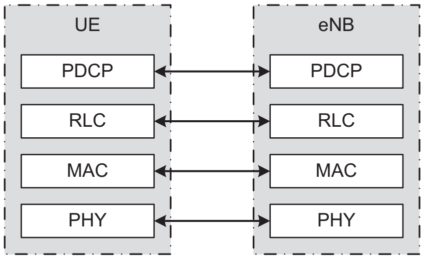


FIGURE 3-20

The user-plane protocol stack [29]

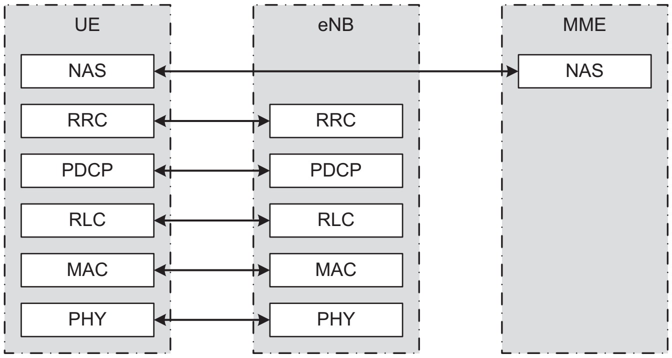


FIGURE 3-21

The control-plane protocol stack [29]

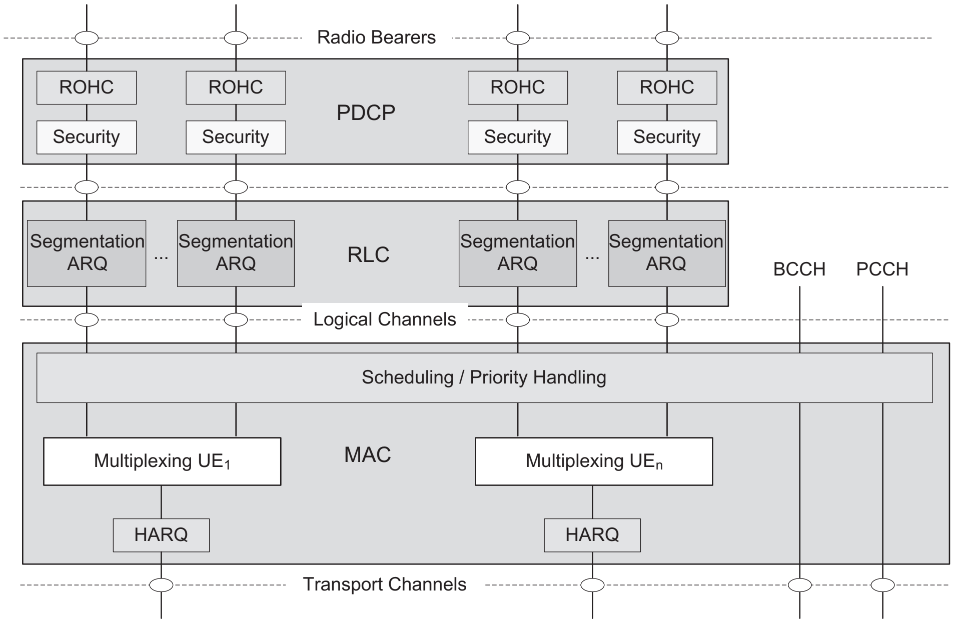


FIGURE 3-22

The 3GPP LTE Layer 2 structure in the downlink [29]

integrity protected by the Mobility Management Entity (MME) and User Equipment (UE), i.e., 3GPP LTE mobile station [28].

The Layer 2 functions in 3GPP LTE are classified into the following categories: Medium Access Control (MAC); Radio Link Control (RLC); and Packet Data Convergence Protocol (PDCP) functions [29]. Figures 3-22 and 3-23 illustrate the structure of Layer 2 protocols in 3GPP LTE downlink and

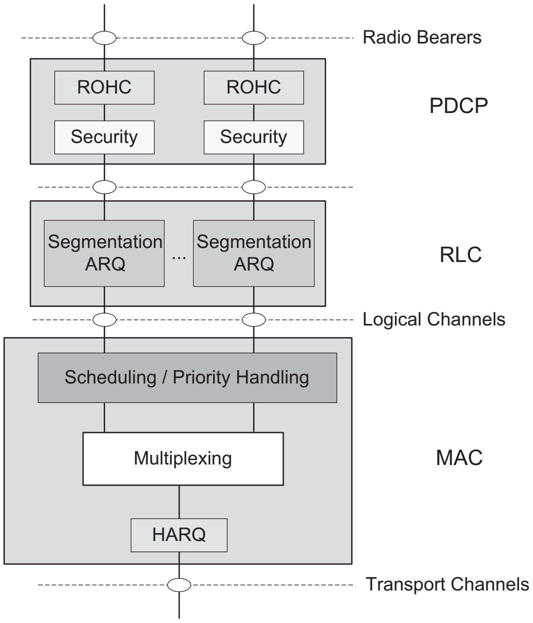


FIGURE 3-23

The 3GPP LTE Layer 2 structure in the uplink [29]

uplink. The SAP for peer-to-peer communication is marked with circles at the interface between the sub-layers. The SAP between the physical layer and the MAC sub-layer provides the transport channels. The SAP between the MAC sub-layer and the RLC sub-layer provide the logical channels. The multiplexing of several logical channels (i.e., radio bearers) on the same transport channel (i.e., transport block) is performed by the MAC sub-layer. Each logical channel is defined by type of information that is transferred. The logical channels are generally classified into two groups: (1) control channels (for the transfer of control-plane information); and (2) traffic channels (for the transfer of user-plane information).

As shown in Figure 3-21, the RRC layer in the evolved NodeB (eNB), i.e., 3GPP LTE base station, makes handover decisions based on neighbor cell measurements reported by the UE, performs paging of the users over the air interface, broadcasts system information, controls UE measurement and reporting functions such as the periodicity of channel quality indicator reports, and further allocates cell-level temporary identifiers to active users. It also executes transfer of UE context from the serving eNB to the target eNB during handover, and performs integrity protection of RRC messages. The RRC layer is responsible for setting up and maintenance of radio bearers. Note that RRC layer in 3GPP protocol hierarchy is considered as Layer 3. The main services and functions of the RRC sub-layer include [29,30]:

- Broadcast of system information;
- Paging;

- Establishment, maintenance, and release of a RRC connection between the UE and E-UTRAN, including allocation of temporary identifiers between UE and E-UTRAN and configuration of signaling radio bearers for RRC connection;
- Security functions, including key management;
- Establishment, configuration, maintenance, and release of point-to-point radio bearers;
- Mobility functions, including UE measurement reporting and control of the reporting for inter-cell and inter-RAT mobility, handover, UE cell selection and reselection, control of cell selection and reselection, context transfer at handover;
- Establishment, configuration, maintenance, and release of radio bearers for Multimedia Broadcast Multicast Service (MBMS);
- QoS management functions.

The 3GPP LTE RRC consists of the following states:

- RRC_IDLE is a state where a UE specific Discontinuous Reception (DRX) may be configured by upper layers. In the idle mode, the UE conserves power and does not inform the network of each cell change. The network knows the location of the UE to the granularity of a few cells, called the Tracking Area (TA). The UE monitors a paging channel to detect incoming traffic, performs neighboring cell measurements and cell selection/reselection, and acquires System Information.
- RRC_CONNECTED is a state where transfer of unicast data to/from UE is performed and the UE may be configured with a UE specific DRX or Discontinuous Transmission (DTX). The UE monitors control channels associated with the shared data channel to determine if data is scheduled for it, provides channel quality and feedback information, performs neighboring cell measurements and measurement reporting, and acquires System Information.

In the U-plane, the PDCP layer is responsible for compressing or decompressing the headers of IP packets using robust header compression to enable efficient use of air interface resources. This layer also performs ciphering of both user-plane and control-plane traffic. Because the NAS messages are carried in RRC, they are effectively double ciphered and integrity protected, once at the MME and again at the eNB. Therefore, the services and functions provided by the PDCP layer in the U-plane include header compression and decompression, transfer of user data between NAS and RLC layer, sequential delivery of upper layer PDUs and duplicate detection of lower layer SDUs at handover for radio link layer acknowledged mode, re-transmission of PDCP SDUs at handover for radio link layer acknowledged mode, and ciphering. The services and functions provided by the PDCP for the C-plane include ciphering and integrity protection, and transfer of control-plane data where PDCP receives PDCP SDUs from RRC and forwards them to the radio link control layer [29,31].

The RLC layer is used to format and transport traffic between the UE and the eNB. The RLC provides three different reliability modes for data transport, i.e., acknowledged mode (AM), unacknowledged mode (UM), and transparent mode (TM). The unacknowledged mode is suitable for transport of real-time services since such services are delay-sensitive and cannot tolerate delay due to re-transmissions. The acknowledged mode is appropriate for non-real-time services such as file transfers. The transparent mode is used when the size of packet data units are known in advance, such as for broadcasting system configuration information. The RLC layer also provides sequential delivery of service data units to the upper layers and eliminates duplicate packets from being delivered to the upper layers. It may also segment the service data units. Furthermore, there are two levels of re-transmissions for providing reliability, the HARQ at the MAC layer, and ARQ at the RLC layer. The

ARQ is required to handle residual errors that are not corrected by HARQ, and is kept simple by the use of a single-bit feedback mechanism. An N-process stop-and-wait HARQ is employed that has asynchronous re-transmissions in the DL and synchronous re-transmissions in the UL. In practice, multiple stop-and-wait HARQ processes are operated in parallel, i.e., when one HARQ process is waiting for an acknowledgment another process can use the channel to send sequentially-ordered sub-packets, thus improving the throughput. The synchronous HARQ means that the re-transmissions of HARQ sub-packets occur at predefined periodic intervals. Hence, no explicit signaling is required to indicate to the receiver the re-transmission schedule. Asynchronous HARQ offers the flexibility of scheduling re-transmissions based on air interface conditions (i.e., scheduling gain) [29,32]. The services and functions provided by the MAC layer can be summarized as follows [32]:

- Mapping between logical channels and transport channels;
- Multiplexing/de-multiplexing of RLC protocol data units corresponding to one or different radio bearers into/from transport blocks delivered to/from the physical layer on transport channels;
- Traffic volume measurement reporting;
- Error correction through HARQ;
- Priority handling between logical channels of one UE;
- Priority handling between UEs through dynamic scheduling;
- Transport format selection.

E-UTRA provides ARQ and HARQ functionalities. The ARQ functionality provides error correction by re-transmissions in acknowledged mode at Layer 2. The HARQ functionality ensures delivery between peer entities at Layer 1. The HARQ within the MAC layer is characterized by an N-process stop-and-wait protocol and re-transmission of transport blocks upon failure of earlier transmissions. A total of eight HARQ processes are supported [29].

The 3GPP LTE-Advanced system extends the capabilities of 3GPP LTE Rel-8 with support of carrier aggregation, where two or more component carriers are aggregated in order to support wider transmission bandwidths up to 100 MHz and for spectrum aggregation. A user terminal may simultaneously receive or transmit one or multiple component carriers depending on its capabilities. From the UE perspective, the Layer 2 aspects of HARQ are similar to those of Rel-8. There is one transport block (in absence of spatial multiplexing, up to two transport blocks in case of spatial multiplexing) and one independent HARQ entity per scheduled component carrier. Each transport block is mapped to

Table 3-1 Summary of the Differences between MAC and RRC Control [33]

	MAC Control		RRC Control
Control Entity	MAC		RRC
Signaling Type	Physical control channel	MAC control PDU	RRC message
Signaling Reliability	$\sim 10^{-2}$ (No HARQ re-transmission)	$\sim 10^{-3}$ (HARQ re-transmissions)	$\sim 10^{-6}$ (ARQ re-transmissions)
Control Latency	Very short	Short	Longer
Extensibility	None	Limited	High
Security	No integrity protection No ciphering	No integrity protection No ciphering	Integrity protected Ciphering

a single component carrier on which all HARQ re-transmissions may take place. A UE may be scheduled over multiple component carriers simultaneously, but at most one random access procedure will be ongoing at any time. Whenever a UE is configured with only one component carrier, the 3GPP LTE Rel-9 DRX is the baseline. In other cases, the same DRX operation will be applied to all configured component carriers. Therefore, the Layer 2 structure of the 3GPP LTE-Advanced is similar to that of 3GPP LTE Rel-8, except for the addition of the multi-carrier functionality; however, the multi-carrier nature of the physical layer is only exposed to the MAC layer through transport channels, where one HARQ entity is required per component carrier [34].

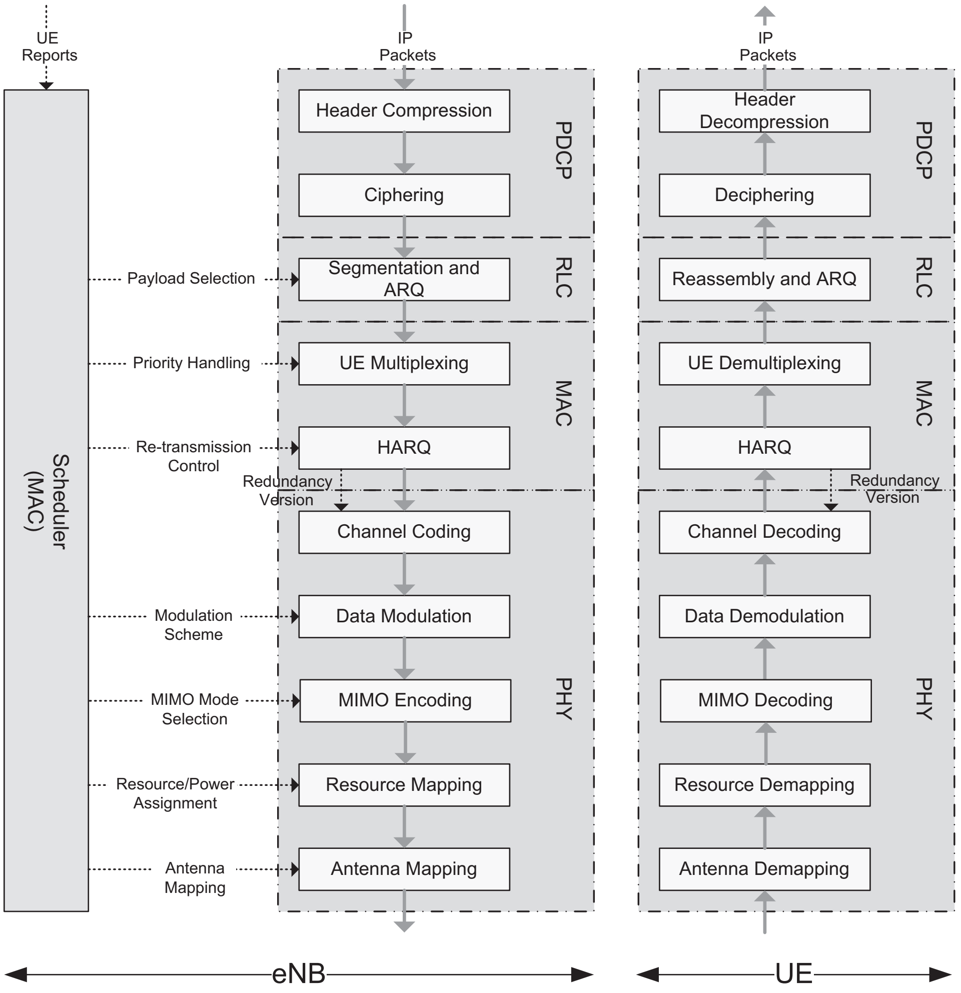


FIGURE 3-24

User data processing and signaling protocols in 3GPP LTE eNB and UE

The main difference between MAC and RRC control lies in the signaling reliability. The signaling corresponding to state transitions and radio bearer configurations should be performed by RRC sub-layer due to signaling reliability. The different characteristics of MAC and RRC control are summarized in Table 3-1.

The physical layer provides information transfer services to MAC and higher layers. The physical layer transport services are described by how and with what characteristics data is transferred over the radio interface. This should be clearly separated from the classification of what is transported, which relates to the concept of logical channels at MAC sub-layer.

Figure 3-24 illustrates the IP packet processing in the eNB and UE transmitter and receiver sides, respectively. The main functions of each protocol layer and their interconnections have been identified. One of the noticeable differences with IEEE 802.16m data processing is the location of encryption or ciphering of the user payload. Unlike the IEEE 802.16m, 3GPP LTE encrypts the packets in the PDCP layer before delivering the service data units to the RLC and MAC layers. The base station scheduler generates appropriate control signals for RLC, MAC, and PHY layers based on periodic reports and measurements received from each UE to ensure robustness and reliability of the connections, given varying radio channel conditions.

References

[1] WMF-T23-001/002/003-R015v01, WiMAX Forum Mobile System Profile Specification: Release 1.5, August 2009, <http://www.wimaxforum.org/resources/documents/technical/release>.

[2] Andrew S. Tanenbaum, Computer Networks, fourth ed., Prentice Hall, 2002.

[3] IEEE Std 802.16-2009, IEEE Standard for Local and Metropolitan Area Networks – Part 16: Air Interface for Broadband Wireless Access Systems, May 2009.

[4] Asynchronous Transfer Mode, Wikipedia, <http://en.wikipedia.org/wiki/ATM_(Asynchronous_Transfer_Mode>.

[5] Open Systems Interconnection Reference Model, Wikipedia, <http://en.wikipedia.org/wiki/OSI_model>.

[6] IETF RFC 1155, M. Rose, Structure and Identification of Management Information for TCP/IP-based Internets, May 1990.

[7] Simple Network Management Protocol, Wikipedia, <http://en.wikipedia.org/wiki/SNMP>.

[8] IETF RFC 1157, J. Case, A Simple Network Management Protocol (SNMP), May 1990.

[9] Common Object Requesting Broker Architecture (CORBA), Wikipedia <http://en.wikipedia.org/wiki/CORBA_architecture>.

[10] Interactive Data Language (IDL), Wikipedia, <http://en.wikipedia.org/wiki/IDL_(programming_language)>.

[11] Web Service, Wikipedia, <http://en.wikipedia.org/wiki/Web_Services>.

[12] IEEE 802.16m–08/003r9, IEEE 802.16m System Description Document, May 2009, <http://ieee802.org/16/tgm/index.html>.

[13] IETF RFC 3418, R. Presuhn, Management Information Base (MIB) for the Simple Network Management Protocol (SNMP), December 2002.

[14] Loutfi Nuaymi, WiMAX: Technology for Broadband Wireless Access, first ed., John Wiley & Sons, 2007.

[15] IETF RFC 791, Internet Protocol, DARPA Internet Program Protocol Specification, September 1981.

[16] IETF RFC 768, J. Postel, User Datagram Protocol, August 1980.

[17] IETF RFC 3550, H. Schulzrinne, RTP: A Transport Protocol for Real-Time Application, July 2003.

[18] Colin Perkins, RTP: Audio and Video for the Internet, first ed., Addison-Wesley, 2003.

[19] IETF RFC 2460, S. Deering, Internet Protocol, Version 6 (IPv6) Specification, December 1998.

[20] IETF RFC 3267, J. Sjoberg, Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs, June 2002.

[21] IETF RFC 3095, C. Bormann, RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed, July 2001.

[22] WiMAX Forum Network Architecture Release 1.5 Version 1 –Stage 2, Architecture Tenets, Reference Model and Reference Points, November 2009, <http://www.wimaxforum.org/resources/documents/technical/release>.

[23] IEEE Std 802.21-2008, IEEE Standard for Local and Metropolitan Area Networks – Part 21: Media Independent Handover Services, January 2009.

[24] Extensible Authentication Protocol (EAP), Wikipedia, <http://en.wikipedia.org/wiki/Extensible_Authentication_Protocol>.

[25] Femto Forum, <http://femtoforum.org>.

[26] Femtocell, Wikipedia, <http://en.wikipedia.org/wiki/Femto_cell>.

[27] IEEE Std 802.16j-2009, IEEE Standard for Local and Metropolitan Area Networks Part 16: Air-interface for Fixed and Mobile Broadband Wireless Access Systems, Multi-hop Relay Specification, July 2009.

[28] E. Dahlman, et al., 3G Evolution: HSPA and LTE for Mobile Broadband, second ed., Academic Press, October 2008.

[29] 3GPP TS 36.300, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, March 2010.

[30] 3GPP TS 36.331, Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification, March 2010.

[31] 3GPP TS 36.323, Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification, March 2010.

[32] 3GPP TS 36.321, Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification, March 2010.

[33] 3GPP TS 36.322, Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) protocol specification, March 2010.

[34] 3GPP TR 36.912, Feasibility study for Further Advancements for E-UTRA (LTE-Advanced), March 2010.