An Analysis of the Design and Implementation of QoS over IEEE 802.16

Mark C. Wood

Abstract

This paper gives an overview of the QoS mechanisms built into the IEEE 802.16 standard (also known by the name of its vendor interoperability organization, WiMAX). The PHY and MAC layers of 802.16 are described in detail with regards to their QoS aspects. And the relations and interactions of these QoS mechanisms are described to give an understanding of how QoS can be achieved over 802.16. This paper will also provide some factors that make 802.16 distinct in its QoS methods. A comparison of the QoS mechanisms in competing technologies (i.e. 802.11, 3G) will be given. Also, this paper will explore some of the current research being done with 802.16 on developing its QoS performance.

Keywords: IEEE 802.16, WiMAX, QoS, BWA, WMAN, WirelessMAN, multimedia networking

Table of Contents

1.0 Introduction

1.1 Background

Wireless networking has the potential for use in broadband Internet services, video and audio streaming, and as an alternative to the PSTN for voice service. WiMAX phones and PDAs are already being developed by companies like Motorola and Samsung. And, they are already in use in Korea, with WiMAX's cousin technology, WiBro. Soon WiMAX adapters will be built into most laptops. You will not have trouble finding a hot spot. With WiMAX coverage, everywhere will be hot.

There has been some struggle to popularize Broadband Wireless Access (BWA). Most of the delays in the use of BWA up to this point have been caused by legal or political obstructions in the US. Wireless technology is being introduced rapidly in places like China and India that lack the copper wire and fiber cable infrastructures.

However, there have been some legitimate concerns over the use of wireless in broadband networking from a technical standpoint. Some of the primary concerns have been packet loss, atmospheric interference, and contention with other wireless services. The 802.16 standard addresses these concerns with QoS. In short, applying more bandwidth to the right channels at the right time can reduce latency and jitter [Ohrtman05].

1.2 IEEE 802.16 History and Future Potential

The IEEE 802.16 standard, which includes MAC and PHY layer specifications, was developed for Internet services over wireless metropolitan area networks (WMANS), and as an alternative to traditional wired networks, such as DSL and cable-modem. 802.16 has two modes: point-to-multipoint (PMP) mode and mesh mode [Cao05].

One of things that distinguishes 802.16 in the PHY layer is Orthogonal Frequency Division Multiplexing (OFDM), which is a multicarrier modulation scheme. The 802.16 standard has two OFDM-based modes: OFDM and Orthogonal Frequency Division Multiplexing Access (OFDMA). Both of these technologies allow subcarriers to be adaptively modulated (QPSK, 16-QAM, 64-QAM), depending on distance and noise. OFDMA also has scalability options that provide higher efficiency in bandwidth use [Intel04].

The MAC layer of 802.16 was originally designed for point-to-multipoint (PMP) broadband wireless access applications, but later amendments of 802.16a and 802.16d also allow for mesh network architecture [Ghosh05]. And, newly adopted in 802.16a was OFDM, which provides greater spectral efficiency and better mitigation of interference. 802.16b covers most of the QoS aspects of the standard. 802.16TGe introduced scalable OFDMA into the standard. 802.16-2004 revised and replaced previous versions and completed the essential fixed wireless standard. And 802.16e was an enhancement for supporting mobile communications; a handoff mechanism was added, making it possible to support mobile communications at vehicular speeds. A summary of the history of the 802.16 standard, with regard to the QoS aspects of it, is shown in Table 1.

Date	IEEE Standard	Description
Dec 2001	802.16	First standardization for LOS, PMP broadband wireless access applications using 10-66 GHz spectrum
Jan 2003	802.16a	An amendment for NLOS using 2-11GHz spectrum. Allows both PMP and mesh network architecture. OFDM is adopted.
c2003	802.16b	QoS provisioning
Oct 2004	802.16-2004	Revised and replaced previous versions and completed the essential fixed wireless standard
Dec 2005	802.16e	Enhancements to support mobility

Table 1: History of IEEE 802.16 standards, QoS aspects

The WiMAX forum (http://www.wimaxforum.org/home) was formed in June of 2001 to ensure interoperability among products by different vendors. And the IEEE 802.16 Working Group on Broadband Wireless Access Standards (www.wirelessman.org/) is the group developing the IEEE 802.16 WirelessMAN Standard for WMANs. These groups, and a few others, may help popularize WiMAX by bringing vendors together on it. As Vint Cerf, TCP/IP co-developer and Internet pioneer, said "People often take the view that standardization is the enemy of creativity. But I think that standards help make creativity possible -- by allowing for the establishment of an infrastructure, which then leads to enormous entrepreneurialism, creativity, and competitiveness." [Fast Company, April 2000]

A separate technology known as softswitch offers a technology bypass of the PSTN switches. And, 802.16 has a non line-of-site (NLOS) range of 4 miles. The combination of softswitch and 802.16 as a wireless "last mile" makes a viable alternative to the PSTN for voice services. With another technology known as TVoIP, 802.16 can also serve as a cable or satellite TV bypass. And, with the 802.16 standard's ability to transmit 72 Mbps over 30 miles point-to-point, it can also serve as a backhaul bypass. A strong business case can be made for using WiMAX as an alternative to the telephone company's T1 and DS3 data circuits to save the enterprise the monthly local loop charges [Ohrtman05]. And the combination of throughput, power, range and versatility give 802.16 some technological advantages over competing technologies, such as 802.11 and 3G. There are a number of reasons, both economical and technical, that make it an appealing choice for broadband wireless access, but the focus of this paper will be on the QoS aspects.

Following this introduction, Section 2 describes the general QoS mechanisms that are used in the PHY layer, and the ones that are specific to 802.16. Section 3 describes how QoS provisioning is done in 802.16. Section 4 describes the 802.16 MAC layer, with emphasis on the QoS aspects of it. Section 5 describes the bandwidth request-grant mechanism used by 802.16 to provide QoS. Section 6 compares the QoS capabilities of 802.16 with the competing technologies. And, Section 7 explores the current research being done in the field of QoS over 802.16.

Back to Table of Contents

2.0 IEEE 802.16 QoS - General

2.1 The Need for QoS

Like ATM, the 802.16 standard was designed with variety of traffic types in mind. 802.16 has to handle the requirements of very-high-data-rate applications, such as voice over IP (VoIP) and video or audio streaming, as well as low-data-rate applications, such as web surfing, and handle extremely bursty traffic over the

Internet. And it may need to handle all of these at the same time.

Some network applications simply cannot work without QoS. Some delay may be acceptable, but too much can make the application unusable. For example, the IEEE 802.16 group determined that an acceptable delay for VoIP is 120 ms, and over 150 ms delay results in noticeably impaired voice quality. Humans are intolerant of speech delays of over 200 ms. [Hummelholm06]

The signaling and bandwidth allocation algorithms in 802.16 have been designed to accommodate hundreds of connections per channel and allow a variety of QoS requirements. The end user applications may be varied in their bandwidth and latency requirements, so 802.16 must be flexible and efficient over a range of different traffic models [Ghosh05].

2.2 IEEE 802.16 QoS Mechanisms

The 802.16 standard includes several QoS mechanisms at the PHY layer, such as Time Division Duplex (TDD), Frequency Division Duplex (FDD) and Orthogonal Frequency Division Multiplexing (OFDM). Each can help in providing QoS. TDD can dynamically allocate uplink and downlink bandwidth, depending on their requirements. For example, when there is more uplink traffic, more bandwidth can be allocated to that, and when there is less uplink traffic it can be taken away. This is illustrated in Figure 1. Each 802.16 TDD frame is one downlink subframe and one uplink subframe, separated by a guard slot. 802.16 adaptiveley allocates the number of slots for each, depending on their bandwidth needs.

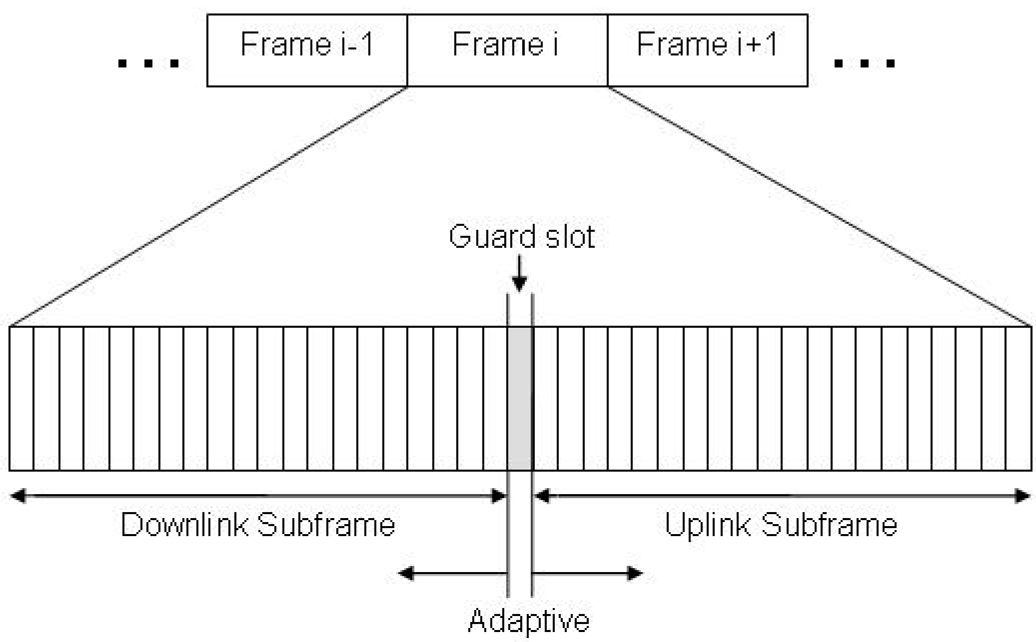

Figure 1: IEEE 802.16 TDD Frame Structure [Maheshwari05]

In FDD, base stations transmit on different sub-bands, and thus do not interfere with each other. This allows for even more bandwidth allocation flexibility. And, OFDM provides greater spectral efficiency, and mitigates interference with its tighter beam width and its dispersal of data across different frequencies [Ohrtman05].

There are a couple of QoS aspects that are specific to OFDM. Forward Error Correction (FEC) builds redundancy into the transmission by repeating some of the information bits, so bits that are missing or in error can be corrected at the receiving end. Without FEC, error correction would require whole frames to be retransmitted, resulting in latency and lower QoS. The other way OFDM helps with QoS is with interleaving. Since OFDM uses multiple subcarriers, a portion of each information bit can be carried on a number of subcarriers, so that if any of the subcarriers is weakened, the information bit can still be received at the destination. And, interleaving means that bits that were close together in time are transmitted on subcarriers that are spaced out in frequency. Thus, errors created on weakened subcarriers are broken up into small bursts and spread out. And these small errors can more easily be corrected with FEC.

OFDM also uses Fast Fourier Transforms (FFT), which in mathematical terms makes evaluating complex numbers more efficient by greatly reducing the number of arithmetical operations required. In radio technology, digital data signals in the form of square waves can be expressed as the sum of a series of sine waves in Fourier's Theorem. OFDM converts a single data stream into M streams and modulates them onto M subcarriers using FFT. In general, FFT can make the transmission of digital signals over the airlink more efficient, which also helps in providing QoS [Nichols01].

The QoS mechanisms described above are some of the general mechanisms that 802.16 uses to ensure good QoS. These mechanisms are already well established in the wireless technology industry, and they have been proven to reduce latency, jitter, and packet loss, which are all goals of QoS.

The 802.16 standard also provides adaptive modulation, which enables dynamic bandwidth allocation to match the current channel conditions. The three modulation schemes used in 802.16 are (from high to low modulation order): 64-QAM, 16-QAM, and QPSK. A higher order modulation scheme can deliver higher throughput, but a higher number of bits per symbol makes it more susceptible to interference and noise. There is a tradeoff between throughput and range. The modulation order can be based on the distance from BS to SS. With adaptive modulation, the BS can adaptively change the code if the distance and atmospheric conditions require it. For example, if the BS cannot establish a robust link to a distant SS using 64-QAM, the modulation order can be reduced to 16-QAM or QPSK. The greater the distance, the lower the QoS guarantee [Ohrtman05].

To summarize, the 802.16 standard incorporates a number of mechanisms to provide QoS as follows:

- Adaptive Modulation (QPSK to QAM 16 to QAM 64)
- Frequency Division Duplex (FDD)
- Time Division Duplex (TDD)
- Orthogonal Frequency Division Multiplexing (OFDM)
- Fast Fourier Transform (FFT)
- Forward Error Correction (FEC)

Every 802.16 implementation will use some combination of these mechanisms to achieve QoS. They are all implemented in the PHY layer, and their parameters are based on the QoS requirements handed down by the higher layers. This is done through QoS provisioning.

2.3 Adaptive Burst Profiles

Another QoS mechanism provided in the PHY level is adaptive burst profiles. Both TDD and FDD configurations support adaptive burst profiling, in which the modulation and coding schemes are specified in a burst profile, where they can be adjusted individually to each SS on a frame-by-frame basis. The burst file allows the modulation and coding schemes to be adaptively adjusted according to link conditions [IEEE 802.16-2004].

Burst profiles describe the UL or DL transmission properties. For UL, the SS transmits in a given time slot with a specific burst size. For DL, multiple SS's can be aggregated in the same DL burst. The burst file parameters allowed are shown below.

Burst profile parameters:

- Modulation Type (QPSK, QAM 16, QAM 64)
- Forward Error Correction (FEC) type
- Preamble length
- Guard times

Different combinations of modulation and FEC provide different QoS assurances. For example, a high order modulation scheme provides a high transmission speed, but is more susceptible to interference. A low order modulation scheme allows a greater distance between BS and SS, but will have a lower transmission speed. FEC can reduce latency by cutting down the retransmissions, but the coding requires more bits, so there is a price to pay if there are no errors. Adaptive burst profiles allow the service requirements to be met under a variety of conditions.

Back to Table of Contents

3.0 IEEE 802.16 QoS Provisioning

The 802.16 standard has three main methods for QoS provisioning:

- Service Flow Classification
- Dynamic Service Establishment
- Two-Phase Activation Model

3.1 Service Flow Classification

The main feature of 802.16 QoS provisioning, and what distinguishes it from its competitors (i.e. 802.11 and 3G), is that it associates each packet with a service flow. 802.16 is a connection-oriented MAC. Each connection is assigned a unique Connection ID (CID) and a Service Flow ID (SFID) with an associated service class. The upper part of the MAC maps data into a QoS service class. Also, external applications can request service flows with desired QoS parameters using a named service class.

The MAC contains a queuing and traffic-shaping engine that is ultimately responsible for the reception and transmission of 802.16a packets according to the enforced QoS parameters. These parameters can be different from service flow to service flow [Redline Communications, http://www.redlinecommunications.com].

802.16 provides four scheduling services, each with an associated service class: UGS, rtPS, nrtPS and BE. These are detailed in Table 2.

QoS Service Class	Description
Unsolicited Grant Service (UGS)	Supports CBR services, such as T1/E1 emulation and VoIP without silence suppression
Real-Time Polling Service (rtPS)	Supports real-time services with variable size data on a periodic basis, such as MPEG and VoIP with silence suppression
Non-Real-Time Polling Service (nrtPS)	Supports non-real-time services that require variable size data grant bursts on a regular basis, such as FTP
Best Effort (BE)	For applications that do not require QoS, such as web surfing

Table 2: IEEE 802.16 QoS Service Classes

Each network application has to register with the network, where it will be assigned one of these service flow classifications with a Service Flow ID (SFID). QoS mapping in the form of classification of higher layer data is provided in the upper part of the MAC. When the application wants to send data packets, the service flow is mapped to a connection using a unique CID [Ganz04].

3.2 Dynamic Service Establishment

The 802.16 standard provides a signaling function for dynamically establishing service flows and requesting QoS parameters. There are three types of control messages for service flows:

- Dynamic Service Activate (DSA) – Activate a service flow
- Dynamic Service Change (DSC) – Change an existing service flow
- Dynamic Service Delete (DSD) – Delete a service flow

New connections may be established when a customer's needs change. This may be initiated by the BS. The BS sends a control message called a DSA-REQ, which can contain the SFID, CID, and a QoS parameter set. The SS then sends a DSA-RSP message to accept or reject the service flow.

This mechanism allows an application to acquire more resources when required. Multiple service flows can be allocated to the same application, so more service flows can be added if needed to provide good QoS.

3.3 Two-Phase Activation Model

Activation of a service flow proceeds in two phases: Admit first, then activate. An Authorization Module in the BS provides this function. It approves or rejects a request regarding a service flow. The Authorization Module can activate a service flow immediately or defer activation to a later time [Ganz04].

Once the service flow has been admitted, both the BS and SS can reserve resources for it. Resources reserved by the BS and SS are not limited to bandwidth. They can include other resources, such as memory. Dynamic changes to the QoS parameters of an existing service flow are also approved by the Authorization Module. QoS parameter changes are requested with Dynamic Service Flow messages sent between the BS and SS. All requests are in the form described above (i.e. DSA, DSC, DSD).

A QoS parameter set is associated with each service flow. The type of QoS parameter set distinguishes the status granted by the Authorization Module (admitted or active). Also, an external set of QoS parameters can be provisioned to the MAC without the Authorization Module, which means they are outside of the scope of the standard. For example, a network management system may provision the QoS parameter set. Altogether, the standard defines three types of QoS parameter sets:

QoS Parameter Set	Description
ProvisionedQoSParamSet	A set of external QoS parameters provided to the MAC, for example by the network management system
AdmittedQoSParamSet	A set of QoS parameters for which the BS and possibly the SS are reserving resources, since the associated service flows have been admitted by the BS
ActiveQoSParamSet	A set of QoS parameters that reflect the actual service being provided to the associated service flow

Table 3: IEEE 802.16 QoS Parameter Sets [Ganz04]

The standard defines an "envelope" model that limits the possible values of the QoS parameter set based on the status of the service flow. An ActiveQoSParamSet is always a subset of the AdmittedQoSParamSet, and the AdmittedQoSParamSet may be a subset of the ProvisionedQoSParamSet. The method for determining which QoS parameters will be allowed depends on the authorization model.

The 802.16 standard recognizes two authorization models:

- Provisioned Authorization: QoS parameters are provided by the network management system upon setup, and remain static
- Dynamic Authorization: Changes to QoS parameters can be requested, and the Authorization Module issues its decisions

Thus, the 802.16 standard provides some flexibility in its QoS provisioning. The two-phase activation allows service flows to be admitted before activation. QoS parameters can also be provisioned by a network management system for use at a later time. And the standard recognizes two authorization models, which allows the vendors some flexibility in their design. This provides an efficient framework for dynamic resource allocation.

The QoS requirements are determined by the higher layer application. For example, a VoIP application may require a real-time service flow with fixed-size data grants, whereas an FTP application may use a non-real-time service flow with variable-sized data grants. If the application requires QoS, it can define the QoS parameter set, or it can imply a set of QoS parameters with a Service Class Name. Depending on the available network resources, the network then decides if it can meet the QoS requirements of the application. If so, the QoS parameters are handed down through the MAC layer.

Back to Table of Contents

4.0 IEEE 802.16 MAC Layer

802.16 is a connection-oriented MAC in that it assigns traffic to a service flow and maps it to MAC connection using a CID. In this way, even connectionless protocols, such as IP and UDP, are transformed into connection-oriented service flows. The connection can represent an individual application or a group of applications sending with the same CID [Wongthavarawat03]. And, the service classes defined in 802.16 are ATM-compatible. Internetworking with ATM is important due to its role in telecom carrier infrastructure and its common use in DSL services.

The MAC layer of 802.16 is divided into two sublayers: the convergence sublayer and the common part sublayer. The convergence sublayer maps the transport-layer-specific traffic into the core MAC common part sublayer. As the name implies, the convergence sublayer handles the convergence of ATM cells and IP packets, so the MAC layer can support both ATM services and packet services, such as IPv4, IPv6, Ethernet, and VLAN services. The common part sublayer is independent of the transport mechanism, and is responsible for fragmentation and segmentation of the SDUs into MAC protocol data units (PDUs), QoS control, and scheduling and retransmission of MAC PDUs [Ghosh05].

The convergence sublayer classifies the incoming Service Data Units (SDUs) by their type of traffic (voice, web surfing, ATM CBR, …) and assigns them to a service flow using a 32-bit SFID. When the service flow is admitted or active, it is mapped to a MAC connection that can handle its QoS requirements using a unique 16-bit CID. A service flow is characterized by a QoS Parameter Set which describes its latency, jitter and throughput assurances. And with Adaptive Burst Profiling, each service flow is assigned a PHY layer configuration (i.e. modulation scheme, Forward Error Correction scheme, etc…) to handle the service.

Once the service flow is assigned a CID, it is forwarded to the appropriate queue. Uplink packet scheduling is done by the BS through signaling to the SS. At the SS, the packet scheduler will retrieve the packets from the queues and transmit them to the network in the appropriate time slots as defined by the Uplink Map Message (UL-MAP) sent by the BS [Wongthavarawat03]. This is illustrated in Figure 2. The IEEE 802.16 QoS architecture can handle multiple levels of QoS through its classification, queuing, and control signaling mechanisms. Figure 2 also illustrates the areas of the architecture that are out of scope of the 802.16 standard and are thus undefined.

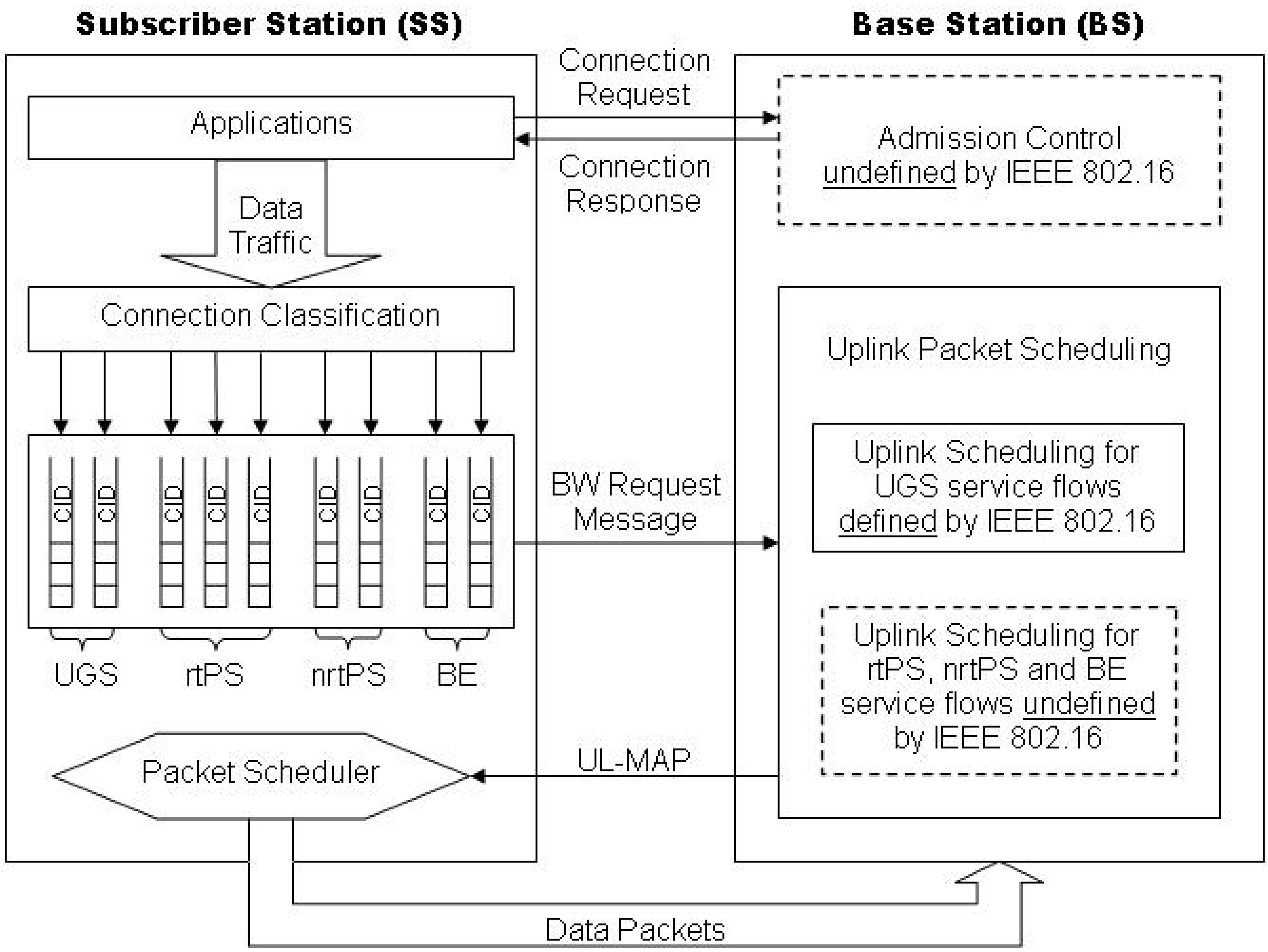

Figure 2: IEEE 802.16 QoS Architecture

Packet header suppression is used to avoid the transmission of redundant information over the airlink. This helps reduce packet delay, which is essential for maintaining an acceptable delay for applications like VoIP. Once the service flow is classified and the CID is assigned to it, the non-changing header information (such as ATM cell header or IP header) is suppressed [Intel04].

The common part sublayer is independent of the transport mechanism. It performs the fragmentation and segmentation of MAC service data units (SDUs) into MAC protocol data units (PDUs). MAC PDUs can be concatenated into bursts having the same modulation and coding. The scheduling and retransmission of MAC PDUs is done in this sublayer. The common part sublayer also performs QoS control. The control signaling for the bandwidth request and grant mechanisms is performed in this sublayer.

Back to Table of Contents

5.0 Bandwidth Requests and Grants

The 802.16 standard provides request and grant mechanisms for bandwidth allocation. These mechanisms enable multiple types of service flows to support a wide range of applications.

5.1 Bandwidth Requests

Bandwidth requests are always per connection. Each connection in an SS requests bandwidth with a BW Request message, which can be sent as a stand-alone packet or piggybacked with another packet. For UGS service flows, the SS does not need to send bandwidth requests because the BS grants it unsolicited.

A bandwidth request can be incremental or aggregate. An incremental bandwidth request means the SS asks for more bandwidth for a connection. An aggregate bandwidth request means the SS specifies how much total bandwidth is needed for a connection. Most requests are incremental, but aggregate requests are occasionally used so the BS can efficiently correct its perception of the SS's needs.

The standard defines several kinds of requests:

- Implicit Requests (for UGS):
  - The SS can request bandwidth implicitly for UGS. In this case, it is negotiated at connection setup, so no actual messages are used.
- BW Request messages
  - Uses a BW Request header
  - Requests up to 32 KB with a single message
  - Incremental or aggregate, as indicated by the MAC header
- Piggybacked request (for non-UGS services only)
  - Presented in GM sub-header and always incremental
  - Up to 32 KB per request for the connection
- Poll-Me Bit
  - Stations request that the BS poll them

Thus, requests can be done in a number of ways. The SS can request bandwidth implicitly (for UGS), which is initiated by the connection. Or, the SS can request bandwidth explicitly in response to receiving a polling message from the BS.

The polling mechanism is detailed and flexible, but requires some overhead. Real-time applications, such as VoIP or video streaming, require a real-time polling service that offers periodic dedicated request opportunities. Non-real-time application, such as FTP, can use a non-real-time polling service that allows random transmit opportunities for bandwidth requests.

The standard allows the BS to poll stations individually or in groups. In total, there are three polling methods:

- Unicast polls: to check for inactive stations
- Multicast and broadcast polls: when there is insufficient bandwidth to poll the stations individually
- Station initiated polls: stations request that the BS poll them (also called the "Poll-Me Bit")

A Best Effort (BE) service is also defined for applications not requiring QoS support. In this case, the SS issues its requests in a contention period.

5.2 Bandwidth Grants

The IEEE 802.16 standard defines two modes for deploying bandwidth grants: Grant Per Connection (GPC) and Grant Per Subscriber Station (GPSS). The BS can use either of these two modes for bandwidth allocation. They are defined as follows:

- GPC
  - BS grants bandwidth explicitly for a connection
  - SS uses it only for that connection
- GPSS
  - BS grants bandwidth to an SS as an aggregate of grants in response to per connection requests from the SS
  - SS may redistribute bandwidth among its connections, maintaining QoS and service-level agreements

GPC is more suitable for few users per subscriber station. It has higher overhead, but allows a simpler SS [Marks05]. GPSS is more suitable for many connections per terminal. It is more scalable, and it reacts more quickly to QoS needs. It has low overhead, but it requires an intelligent SS. GPSS is mandatory for the 10-66 GHz PHY.

Both GPC and GPSS use a self-correcting protocol, rather than an acknowledged protocol. All errors are handled in the same way. After a timeout, the SS simply requests again. This method eliminates the overhead of using acknowledgements, and it uses less bandwidth.

Grants, which are given as durations, are carried in the UL-MAP messages. Referring back to Figure 2, the UL-MAP messages are sent from the Uplink Packet Scheduling system of the BS to the Packet Scheduler of the SS. The BS allocates bandwidth to the connection or SS by granting the transmission opportunity in the UL-MAP. The decision to grant bandwidth is mainly based on the amount requested, the needs of the current service flows, and the available network resources.

QoS Service Type	Implementation
Unsolicited Grant Service (UGS)	BS provides fixed-size data grant bursts periodically
Real-Time Polling Service (rtPS)	BS provides SS the opportunity to request bandwidth on a regular basis
Non-Real-Time Polling Service (nrtPS)	BS provides SS opportunity to request bandwidth using unicast and contention methods
Best Effort (BE)	BS allows SS to use all available mechanisms for transmission requests

Table 4: QoS Service Class Implementation [Ahmad05]

The four service classes, described previously in Table 2, are implemented by using the QoS mechanisms built into the grant-based MAC. Table 4 describes how each service class is implemented. They are achieved using the bandwidth request-grant mechanisms described above, and a combination of all the other QoS mechanisms are used to ensure good QoS.

Back to Table of Contents

6.0 Comparison with Competitor's QoS

6.1 IEEE 802.11

The 802.11a revision endowed 802.11 with QoS capabilities, allowing prioritization of different traffic types and contention-free transmissions over a short period of time. But since 802.11 remains Ethernet-based, instead of IP-based, there are a number of ancillary QoS protocols not available [Sweeny06]. One disadvantage is that it is a connectionless protocol; the MAC uses acknowledgements which cause overhead and delays. And the MAC is contention based, which means acknowledgements and timeouts may cause delays if there are errors. And 802.11 uses time slots, so there is no pre-emption. Also, 802.11 allows bursting, which can be a problem with Internet service. In 802.11, the channel size is fixed, unlike 802.16 which has a channel size that is changeable. The main problem is that 802.11 lacks throughput, power, and range, which make QoS guarantees difficult. 802.11 did not offer "the last mile". And, it is inefficient. The extra frames used in 802.11 could be considered a waste of time. In 802.16, no time is wasted between users (i.e. There are no DIFS, SIFS, PIFS, etc…).

6.2 3G

3G is still not an "all IP" solution, but that is the goal. With 3G, IP QoS control must be mapped onto the circuit-switching layer, which leads to corruption. And the mapping point in the core can be far away from the delivery point, creating queuing and scheduling inefficiencies. In 3G, things are generally fixed. There is no adaptive modulation. There is HSDPA technology which includes Hybrid ARQ (HARQ) to allow it to dynamically adjust to network conditions somewhat, but it does not provide as much flexibility as OFDM and adaptive modulation. 3G may be thought of as a cellular network, which is a different industry. Things are more competitive in the data world. In 3G, the cost is high for data networking, so the cellular companies would generally rather use the bandwidth to sell voice service. 802.16 has a clear advantage over 3G in price.

6.3 Advantages of QoS over IEEE 802.16

802.16 is a connection-oriented protocol. It uses service flows, which is a fundamental distinction from connectionless protocols, like 802.11. And, 802.16 provides for ATM-compatible QoS service classes. 802.16 is a grant-based MAC, as opposed to contention-based like 802.11. A grant-based MAC allows centralized control and eliminates the overhead and delay of acknowledgements, which in turn allows better QoS handling. And, the SS can react to QoS needs in real time. It also reduces the workload of the BS. And there is lower overhead since connections are aggregated. (One disadvantage is that it requires an intelligent SS to redistribute the allocated bandwidth among connections.)

The fragmentation of PDUs allows very large SDUs to be sent across frame boundaries to guarantee QoS of competing services. OFDM provides error correction and interleaving to improve QoS. And adaptive modulation corrects distance problems. 802.16 allows for channel divisions based upon frequency divisions or time slots. And it provides 1.5 to 28 MHz channels, where the channel size is changeable (In 802.11, it is fixed.). And, 802.16 supports a mesh mode which, unlike traditional cellular systems, the nodes can communicate without having a direct connection with the base station [Shetiya05].

Technology	QoS Advantages/Disadvantages
IEEE 802.11	Contention-based MAC, requires acknowledgements, which causes overhead, latency, timeouts; Uses time slots, no pre-emption; Fixed channel size
3G	Still not an "all IP" solution; IP QoS must be mapped onto circuit-switching layer, leads to corruption; Mapping point may be far away, causing queuing and scheduling inefficiencies; Most parameters are fixed, not adaptive
IEEE 802.16	Connection-oriented protocol, provides service flows; Grant-based MAC allows centralized control and eliminates overhead and delay of acknowledgments; Reacts to QoS needs in real time; OFDM, FEC and adaptive modulation for flexible and efficient QoS

Table 5: QoS Advantages and Disadvantages of Competing Technologies

Table 5 summarizes the advantages and disadvantages of these technologies. Only QoS aspects are listed.

Back to Table of Contents

7.0 Suggestions for Improvement

Ongoing research is being done on improving QoS performance in 802.16. Resource management, traffic shaping, and access control in the BS is out of the scope of the 802.16 standard. The 802.16 standard includes a number of QoS signaling mechanisms, but algorithms that use such signaling mechanisms were left out of the standard. This allows vendors to develop their own algorithms for bandwidth allocation and differentiate their own products, but remain interoperable with their core design [Ganz04].

In the area of QoS over 802.16, most research is focused is on packet scheduling, with a little research being done on link adaptation and MPLS. Packet scheduling algorithms, which determine the uplink and downlink bandwidth allocation, are undefined. In order to provide different levels of QoS guarantees for various applications while still achieving high system utilization, the QoS architecture should be integrated into the MAC protocol [Chu02]. Packet scheduling is the hottest area of research in 802.16. There are many papers that propose a new architecture for 802.16 with their own packet scheduler.

Figure 2 showed the specific areas in the QoS architecture that are undefined: Admission Control and Uplink Scheduling for rtPS, nrtPS and BE service flows. [Wongthavarawat03] propose an architecture that defines these areas. These papers introduce a scheduling algorithm and admission control policy for 802.16. They also suggest system parameters that may be used, and define traffic characteristics for which the network can provide QoS.

Some other papers that propose a new architecture with packet scheduling are as follows. [Maheshwari05], which was submitted as a Master's thesis, provides a detailed description of the proposed architecture and more background on the 802.16 standard. [Liu05A] presents a scheduler where the priority is based on channel and service quality. [Shang05] proposes a scheduler for the uplink that uses hard QoS and best effort for the BS, and validates fairness. [Liu05B] proposes a scheduling architecture based on WFQ. [Hawa02] presents a scheduling architecture for both 802.16 and Data Over Cable Service Interface Specifications (DOCSIS). Some other proposals for a scheduling architectures are described in [Chu02], [Cho05] and [Moraes05].

In addition to packet scheduling, some other interesting research has been going on. [Ramachandran04], which was submitted as a Master's thesis, focuses on link adaptation (the matching of modulation, coding and other signaling and protocol parameters to conditions on the radio link). [Cao05] focuses on mesh mode and presents a distributed scheduler for 802.16 mesh networks. [Shetiya05] is also for 802.16 mesh networks, but with a centralized scheduling architecture. This paper proposes a scheduling algorithm for applications using UDP for voice and video, and a scheduling algorithm for applications using TCP for data, and combines these algorithms to handle both. [Chen05] proposes a system integrating IntServ for mesh networks, DiffServ for mapping between IP and MAC layers, and a signaling mechanism for PMP, and compares this system with RSVP. [Sethom04] proposes an MPLS architecture that can be applied to different wireless technologies, including 802.16. And, [Gakhar05] proposes an architecture to map 802.11e QoS requirements to 802.16, and proposes a mapping for different service flows.

In summary, most of the research is focused on packet scheduling, but it is not clear yet which scheduling algorithms will work best with 802.16. Most of the testing has been done in simulation, so one of the main challenges may be to get reliable performance evaluations to create a heuristic model. And there are other areas of research in 802.16 that are just getting started. Since 802.16 is new, the vendors do not know the answers yet and are very interested in this research.

Back to Table of Contents

8.0 Summary

There are a number of concerns over broadband wireless access, such as packet loss and atmospheric interference, but these problems can be mitigated with QoS. The 802.16 MAC is connection-oriented and provides Service Flows, each assigned a QoS service class. 802.16 also uses Dynamic Service Establishment and a Two-Phase Activation Model for QoS provisioning. The 802.16 PHY layer includes OFDM, FEC and Adaptive Modulation to improve QoS performance. And Adaptive Burst Profiles allow the modulation and coding schemes to be adaptively adjusted according to link conditions. The 802.16 standard is a grant-based protocol, which provides centralized control and low overhead, as opposed to a contention-based protocol such as 802.11 that relies on acknowledgements. And 802.16 has the power and range to provide the "last mile" in voice services, at a lower cost than 3G. 802.16 research is hot now as people are trying to build the standard and improve its performance. And, QoS over 802.16 may be the key to getting wider acceptance of broadband wireless access for voice and multimedia services.

Back to Table of Contents

References

1. [Ahmad05] Aftab Ahmad, "Wireless and Mobile Data Networks", ISBN: 0471670758, Wiley-Interscience, July 20, 2005, p.319
2. [Cao05] Min Cao, Wenchao Ma, Qian Zhang, Xiaodong Wang, Wenwu Zhu, Modelling and Performance Analysis of the Distributed Scheduler in IEEE 802.16 Mesh Mode, MobiHoc 05, May 2005
3. [Chen05]Jianfeng Chen, Wenhua Jiao, Qian Guo, An Integrated QoS Control Architecture for IEEE 802.16 Broadband Wireless Access Systems, Lucent Technologies, Bell Labs Research China, c2005
4. [Cho05] Dong-Hoon Cho, Jung-Hoon Song, Min-Su Kim, Ki-Jun Han, "An Architecture for Efficient

QoS Support in the IEEE 802.16 Broadband Wireless Access Network", Springer, Berlin / Heidelberg, ISSN: 0302-9743, c2005

5. [Chu02] GuoSong Chu, Deng Wang, Shunliang Mei, A QoS Architecture for the MAC Protocol of IEEE 802.16 BWA System, IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions, June 2002
6. [Eklund02] Carl Eklund, Roger B. Marks, Kenneth L. Stanwood, Stanley Wang, IEEE Standard 802.16: A Technical Overview of the WirelessMAN Air Interface for Broadband Wireless Access, IEEE Communications Magazine, June 2002
7. [Flarion06]"Flarion," http://www.flarion.com, developers of WiMAX technology products, such as the RadioRouter basestation, which uses their "Flash-OFDM" technology (white paper download page located at http://www.flarion.com/products/white_papers.asp)
8. [Gakhar05] Kamal Gakhar, Annie Gravey and Alain Leroy, IROISE: A New QoS Architecture for IEEE 802.16 and IEEE 802.11e Interworking, COMNETS 2005, colocated with BROADNETS 2005, Boston, MA, Oct. 2005
9. [Ganz04] Aura Ganz, Zvi Ganz, Kitti Wongthavarawat, "Multimedia Wireless Networks: Technologies, Standards, and QoS," ISBN: 0130460990, Prentice Hall PTR, c2004, pp.119-139
10. [Ghosh05] A. Ghosh, D. Wolter, J. Andrews, R. Chen, Broadband Wireless Access with WiMax/802.16: Current Performance Benchmarks and Future Potential, IEEE Communications Magazine, vol. 43, no. 2, February 2005, p.131
11. [Hawa02] Mohammed Hawa, David W. Petr, Quality of Service Scheduling in Cable and Broadband Wireless Access Systems, Tenth International Workshop on Quality of Service, pp.247-255, May 2002
12. [Hos04] Eric Hos, Niels Visser, Quality of Service for wireless networks, Analytical Network Project Report, University of Amsterdam, the Netherlands, June 29, 2004
13. [Hoymann05] Christian Hoymann, "Analysis and performance evaluation of the OFDM-based metropolitan area network IEEE 802.16", Computer Networks, Selected Papers from the European Wireless 2004 Conference, Vol. 49, No. 3, pp.341-363, The Hague, Netherlands, October 2005
14. [Hummelholm06] Aarne Hummelholm, Wimax MAC – Upperlayer Services (pdf slides), HUT Communications Laboratory, Finland, March 28, 2006
15. [IEEE 802.16-2004] IEEE 802.16-2004 (802.16REVd) Specification, "802.16-2004 IEEE Standard for Local and Metropolitan Area Networks, Part 16, Air Interface for Fixed Broadband Wireless Access Systems," June 24, 2004, p.31, p.207
16. [IEEE 802.16 Working Group] "The IEEE 802.16 Working Group on Broadband Wireless Access Standards," www.wirelessman.org/, group developing the IEEE 802.16 WirelessMAN Standard for WMANs. Includes technical reports, tutorials, presentations, and links relating to IEEE 802.16.
17. [Intel04] "Intel Technology Journal on WiMAX," Online article at http://download.intel.com/technology/itj/2004/volume08issue03/vol8_iss03.pdf, ISSN 1535-864X, Vol. 8, Issue 3, August 20, 2004, p.215
18. [Liu05A] Qingwen Liu, Xin Wang, Georgios B. Giannakis, "Cross-Layer Scheduler Design with QoS Support forWireless Access Networks", Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, Aug. 22-24, 2005
19. [Liu05B] N. Liu, X. Li, C. Pei, B. Yang, "Delay Character of a Novel Architecture for IEEE 802.16 Systems", Proceedings of the Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'05), 2005
20. [Maheshwari05] Supriya Maheshwari, An Efficient QoS Scheduling Architecture for IEEE 802.16 Wireless MANs, Master's thesis, Indian Institute of Technology, Bombay, c2005
21. [Marks05] Roger Marks, The IEEE 802.16 WirelessMAN Standard for Broadband Wireless Metropolitan Area Networks (pdf slides), IEEE Computer Society Distinguished Visitors Program, Oakland-East Bay, 19 July 2005
22. [Moraes05] L. F. M. de Moraes, P. D. Maciel Jr., Analysis and Evaluation of a New MAC Protocol for Broadband Wireless Access, International Conference on Wireless Networks, Communications, and

Mobile Computing - WirelessCom 2005, Maui, Hawaii, June 2005

23. [Nichols01] Randall K. Nichols, Panos C. Lekkas, “Wireless Security: Models, Threats, and Solutions,” ISBN: 0071380388, McGraw-Hill Professional, December 13, 2001, p.283
24. [Ohrtman05] Frank Ohrtman, “WiMAX Handbook: Building 802.16 Wireless Networks,” ISBN: 1590595742, McGraw-Hill, May 24, 2005, pp.54-75
25. [Parekh06] Shyam Parekh, 802.16/WiMAX (PPT only), University of California, Berkeley, 2006
26. [Ramachandran04] Shyamal Ramachandran, "Link Adaptation Algorithm and Metric for IEEE Standard 802.16", Master's Thesis, Virginia Polytechnic Institute and State University, http://scholar.lib.vt.edu/theses/available/etd-02272004-133201/unrestricted/ramachandran_thesis.pdf, February 12, 2004
27. [Redline06] “Redline,” http://www.redlinecommunications.com/, received WiMAX Forum Certification mark for their product, RedMAX. Site includes many WiMAX resources.
28. [Sethom04] Kaouthar Sethom, Hossam Afifi, Guy Pujolle, "Wireless MPLS: a new layer 2.5 micro-mobility scheme", Pages 64-71, Electronic Edition (ACM DL), Proceedings of the Second International Workshop on Mobility Management & Wireless Access Protocols, 2004, Philadelphia, PA, USA, October 1, 2004. ACM 2004
29. [Shang05] Yanlei Shang, Shiduan Cheng, "An Enhanced Packet Scheduling Algorithm for QoS Support in IEEE 802.16 Wireless Network", Networking and Mobile Computing: 3rd International Conference, ICCNMC 2005, Zhangjiajie, China, August 2-4, 2005, Proceedings, Springer 2005
30. [Shetiya05] H. Shetiya, V. Sharma, Algorithms for Routing and Centralized Scheduling to Provide QoS in IEEE 802.16 Mesh Networks, WMuNeP 2005, Montreal, Quebec, Canada, October 2005
31. [Sweeney06] Daniel Sweeney, “WiMax Operator's Manual: Building 802.16 Wireless Networks,” ISBN: 0071454012, Apress, c2006
32. [WiMAX Forum06] “WiMAX Forum,” www.wimaxforum.org/home, news, events, white papers, links, from the group trying to make WiMAX interoperable (The download page is located at www.wimaxforum.org/news/downloads/).
33. [Wongthavarawat03] Kitti Wongthavarawat, Aura Ganz, Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems, Wiley, c2003

Back to Table of Contents

List of Acronyms

BE	Best Effort
BS	Base Station
CBR	Constant Bit Rate
CID	Connection ID
DOCSIS	Data Over Cable Service Interface Specifications
DSA	Dynamic Service Activate
DSA-REQ	Dynamic Service Activate-Request
DSA-RSP	Dynamic Service Activate-Response
DSC	Dynamic Service Change
DSD	Dynamic Service Delete
FDD	Frequency Division Duplex
FEC	Forward Error Correction

FFT	Fast Fourier Transforms
GPC	Grant per Connection
GPSS	Grant per Subscriber Station
HARQ	Hybrid Automatic Repeat-reQuest
IPv4	IP version 4
IPv6	IP version 6
MPLS	Multiprotocol Label Switching
NLOS	Non Line of Sight
nrtPS	Non Real-time Polling Service
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiplexing Access
PDU	Protocol data unit
PMP	Point-to-multipoint
PSTN	Public Switched Telephone Network
Qos	Quality of service
QPSK	Quadrature Phase Shift Keying
QAM	Quadrature amplitude modulation
rtPS	Real-time Polling Service
SDU	Service Data Unit
SFID	Service Flow ID
TDD	Time Division Duplex
UGS	Unsolicited Grant Services
UL-MAP	Uplink Map Message
VLAN	Virtual LAN
VoIP	voice over IP
WFQ	Waited Fair Queuing
WMANS	Wireless metropolitan area networks

Back to Table of Contents

Last Modified: April 23, 2006

Note: This paper is available online at
http://cec.wustl.edu/~mcw2/QoS_over_802_16/QoS_over_802_16.html