CROSS LAYER ROUTING IN WiMAX MESH NETWORKS

by

Mehmet Şükrü Kuran

B.S., in Computer Engineering, Yıldız Technical University, 2004

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in System and Control Engineering
Boğaziçi University
2007

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my thesis supervisor Assist. Prof. Tuna Tuğcu for allowing me to work with such an insightful and helping professor. Without his detailed comments, genuine ideas, and endless patience this thesis would not come into fruition.

Secondly, I would like to thank my mother, Lale Kuran, and my father Ali Haydar Kuran for their teaching on being an exemplary human being even though I still could not fulfill even a small fragment of. I specifically thank to my mother for her technical ideas from an old engineer to a new engineer and her endless patience.

I would like to specially thank to H. Birkan Yılmaz, Gürkan Gür, and İlker Demirkol for their both insight and support during this thesis. I also would like to thank all current and former fellow Satellite Laboratory (SATLAB) and Network Laboratory (NETLAB) researchers for their support and patience during this thesis. Namely, H. Birkan Yılmaz, Evren Önem, Pınar Santemiz, Neşe Alyüz, Bora Zeytinci, Fernaz Alimoğlu, Suzan Bayhan, Gürkan Gür, Kaan Bür, İlker Demirkol, Rabun Koşar, and B. Atay Özgövde. I apologize for the many inconveniences I have caused to them and their work during my thesis. Last but not least, I would like to thank my friends Burak Pilavcı, Yahya Çolak, Saner Kumbasar, Burak Doğruöz, Cumhut T. Zümrüt, İlker Kopan, and Selin Gürsoy for their morale support before and during this thesis. I also thank Burak Doğruöz, Cumhur T. Zümrüt and İlker Kopan for their logistic support during the last couple of years.

This research is supported by the Scientific and Technical Research Council of Turkey (TÜBİTAK) under grant number 104E032 and State Planning Organization of Turkey under grant number DPT 2003 K120250, "The Next Generation Satellite Networks and Applications" project.

ABSTRACT

CROSS LAYER ROUTING IN WiMAX MESH NETWORKS

The Mesh mode of the WiMAX standard divides the system resources into two regions, one for Internet other for intranet traffic. According to the standard, one type of traffic cannot use the other's reserved capacity in anyway. This results in reduced system utilization when there is congestion in one region while other is not congested. We develop a cross layer Queue Aware Routing (QAR) scheme for the Mesh mode of IEEE 802.16 with two approaches that utilize the capacity allocated to intranet traffic Internet traffic in case the latter suffers from congestion. Our solution does not introduce any changes in the standard. Thus, it can be implemented without incurring any change to existing user and BS devices. Our simulation results show that both of our QAR schemes decrease the end-to-end delay and the number of dropped packets at the cost of a slight increase in the delay for the intranet traffic.

ÖZET

WiMAX MESH AĞLARINDA KATMANLAR ARASI YÖNLENDİRME

WiMAX standardının Mesh çalışma modu sistem kapasitesini Internet ve intranet trafiği için iki bölgeye ayırmıştır. Standarda göre bu iki trafik türünden biri diğerinin kapasitesini hiçbir şekilde kullanamamaktadır. Bu durum iki bölümden birinde sıkışıklık ve diğerinde boş kapasite olduğu durumlarda düşük sistem kullanımına yol açmaktadır. Biz çalışmamızda IEEE 802.16'nin Mesh çalışma modu için katmanlar arası çalışan Kuyruk Durumundan Haberdar bir Yönlendirme metodu geliştirdik. Bu metodu iki farklı yaklaşımla inceledik. Metodumuz, Internet trafik bölgesinde sıkışıklık olduğu durumlarda Internet trafiğinin, intranet bölümünü de kullanmasına olanak sağlamaktadır. Çözümümüz standartta herhangi bir değişiklik gerektirmediğinden mevcut kullanıcı ve baz istasyonu cihazlarında kolaylıkla uygulanabilmektedir. Simulasyon sonuçlarımız metodumuzun iki yaklaşımınında, Internet trafiğinde sıkışıklık meydana geldiği durumlarda uçtan uca gecikmeyi ve düşen paket adedini azalttığını göstermektedir. Bunun yanısıra Internet trafiğinin, intranet trafiğine ayrılan bölgeyi kullanmiş olduğu için metodumuz intranet trafiğinin uçtan uca gecikme değerlerini fazla olmamakta birlikte artırmaktadır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

$CLCth$	Centralized Link Congestion Threshold
$CLUth$	Centralized Link Uncongestion Threshold
$Csto$	The Available Number of TOs in the CS part of the control subframe
$Csto_{max}$	The Maximum Number of TOs in the CS part of the control subframe
$DLCth$	Distributed Link Congestion Threshold
f	Flow Scale Exponent
$GrtFrCnt$	The Number of Frames Grant Messages Take to Reach Every SS in the Topology
$GrtUp_i$	Uplink grant of SS_i(in bits/sec)
gup_i	Uplink grant of SS_i(in 4-bit reduced form)
Hc_i	Hop Count of ith Subscriber Station
Max_{Hc}	The Highest Hop Count in the Network
N	Number of SSs in the Topology
$NodeWChldCnt$	Number of SSs with Child Nodes
$ReqFrCnt$	The Number of Frames Required for Every SS to Send Their Uplink Bandwidth Requests
$ReqUp_i$	Uplink request of SS_i(in bits/sec)
rup_i	Uplink request of SS_i(in 4-bit reduced form)
$QARch_i$	The time (in seconds) SS_i starts using the QAR system
SS_i	ith Subscriber Station

AAS	Adaptive Antenna System
ACK	Acknowledgement
ACQPS	Active QoSParamSet
APSK	Amplitude Phase Shift Keying
AQPS	AdmittedQoSParamSet
ARQ	Automatic Response Request
BE	Best Effort

BER	Bit Error Rate
BM	Basic Management
BPSK	Binary Phase Shift Keying
BWA	Broadband Wireless Access
CBR	Constant Bit Rate
CDC	Combined Distributed and Centralized
CDS	Coordinated Distributed Scheduling
CID	Connection Identifier
CPS	Common Part Sublayer
CQAR	Centralized Queue Aware Routing
CRC	Cyclic Redundancy Check
CS	Convergence Sublayer
CSCF	Centralized Scheduling Configuration
CSCH	Centralized Scheduling
DCD	Downlink Channel Description
DFS	Dynamic Frequency Selection
DIUC	Downlink Interval Usage Code
DLC	Data Link Control
DL-MAP	Downlink Map
DCQAR	Distributed, Centralized Queue Aware Routing
DRR	Deficit Round Robin
DSCH	Distributed Scheduling
ertPS	Extended Real Time Polling Service
ETE	End to End
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FCH	Frame Control Header
FDD	Frequency Division Multiplexing
FEC	Forward Error Correction
HCA	Hop Count Aware
ID	Identifier

IEEE	Institute of Electrical and Electronics Engineers
LMDS	Local Multipoint Distribution System
LOS	Line of Sight
MBS	Mesh Base Station
MBWA	Mobile Broadband Wireless Access
MMDS	Multichannel Multipoint Distribution System
MRR	Mobile Multihop Relay
MS	Mobile Subscriber
MSS	Mesh Subscriber Station
NLOS	Non LOS
nrtPS	non-Real Time Polling Service
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
Per	Persistence
P2P	Point to Point
PDU	Packet Data Unit
PHS	Packet Header Suppression
PKM	Privacy Key Management
PM	Primary Management
PMB	Poll Me Bit
PMP	Point to Multipoint
PQPS	ProvisionedQoSParamSet
PS	Physical Slot
QAM	Quadrature Amplitude Modulation
QAR	Queue Aware Routing
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RS	Relay Station
rtPS	Real Time Polling Service
SA	Security Association
SAID	Security Association Identifier

SAP	Service Access Point
SC	Single Carrier
SDMA	Space Division Multiple Access
SDU	Segment Data Unit
SF	Service Flow
SFID	Service Flow Identifier
SM	Secondary Management
SNR	Signal to Noise Ratio
SR	Slot Range
SS	Subscriber Station
TDD	Time Division Duplexing
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TO	Transmission Opportunity
UCD	Uplink Channel Description
UDS	Uncoordinated Distributed Scheduling
UGS	Unsolicited Grant Service
UIUC	Uplink Interval Usage Code
UL-MAP	Uplink Map
VBR	Variable Bit Rate
VC	Virtual Channel
VCI	Virtual Channel Identifier
VN	Virtual Node
VP	Virtual Path
VPI	Virtual Path Identifier
VoIP	Voice over IP
WFQ	Weighted Fair Queuing
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WPS	Wireless Packet Scheduling

WWAN	Wireless Wide Area Network

1. INTRODUCTION

In recent years, the number and importance of multimedia Internet applications have increased significantly. Unlike conventional Internet applications like Web, file transfer and mail, multimedia applications have high bandwidth requirements. In order to support these services\ various broadband Internet access technologies have been developed. Initially based on wired solutions, these technologies had limited support for nomadic and mobile users. Thus, demand for wireless broadband Internet access has increased. These wireless broadband access technologies are explained in detail in [1].

One of the emerging wireless broadband access technologies, WiMAX, provides Internet access rates comparable to DSL and cable systems. WiMAX has two operation modes: *Point to Multipoint* (*PMP*) and Mesh. All users must be connected to the BS in the PMP mode. On the other hand, in the Mesh mode user stations that cannot connect directly to the BS can relay their signals over other user stations in range. Also, in the Mesh mode users can communicate directly with each other. The allowance of intranet traffic and increased coverage makes the Mesh mode of WiMAX a more versatile than the PMP mode.

Being a broadband access technology most of the traffic generated by a WiMAX user is destined to the Internet. However, some network capacity is allocated also for the intranet traffic. Thus, in the standard the total system capacity is divided into two parts, one for Internet traffic and the other for intranet traffic. While the Internet traffic is managed by the BS, intranet traffic is managed by the *Subscriber Stations* (*SS*)s. This differentiation scheme that is currently in use has one major drawback. In cases where offered Internet traffic exceeds the allocated capacity congestion occurs even if the intranet traffic part is low. Since the BS does not know how much of the intranet part is being used at a given time, it cannot take any action to increase the capacity allocated to the Internet traffic.

WiMAX standard is a data link layer protocol. Thus, it does not specify how the traffic will be routed in the Mesh topology. In this thesis we develop a cross layer *Queue Aware Routing* (*QAR*) mechanism that uses the queue length values of WiMAX links as decision parameters and changes the routing of Internet traffic in congestion scenarios to reduce Internet traffic's ETE packet delay and packet drops. Two different decision parameter sets are used in this work. Both approaches in our mechanism temporarily use the intranet traffic part of the network in such scenarios to carry the excess Internet traffic. Our mechanism strictly abides by the IEEE 802.16 standard and can be implemented without any change in the standard both in BS and user devices. The simulation results show that both decision parameter sets of our routing scheme are effective in Internet traffic congestion scenarios without causing too much hindrance to the intranet traffic.

In the literature, there are various simulation models regarding the PMP mode of WiMAX. However, a detailed system model for the Mesh mode does not exist. In this thesis we develop a system model for the Mesh mode of WiMAX using OPNET 11.5 simulation software. Also, there are some parts in the standard that are left unstandardized intentionally. We also develop algorithms for these parts for our model to work. In addition to our routing scheme, we develop a BS scheduler and the distributed scheduling requester.

The rest of the thesis is as follows. In Chapter II, the IEEE 802.16 ,WiMAX, standard and its mechanisms are explained in detail. Chapter III explains the QoS and scheduling mechanisms of the standard and propositions developed in the literature regarding these mechanisms. In Chapter IV we propose QAR mechanism including decision parameter sets and additional mechanisms. Simulation environment and results are shown in Chapter V. Finally, we conclude the thesis in Chapter VI.

2. IEEE 802.16 / WiMAX

The IEEE 802.16 standard is developed based on two systems; *Multichannel Multipoint Distribution System* (*MMDS*) and *Local Multipoint Distribution System* (*LMDS*). Starting from 1996, some telephony companies started developing propriety wireless broadband access technologies as an alternative to DSL and cable services. These services, called MMDS, provide data rates up to several Mbps. In 1998, FCC allocated frequency bands for these services [4]. In order to provide good service quality in urban settings, 2.1 GHz and 2.5-2.7 GHz frequency bands which are very good against rain and vegetation attenuation are chosen for MMDS. A typical MMDS cell usually has a radius of 50 km and gives 0.5-30 Mbps aggregate data rate per cell [5].

Due to its ease of deployment, MMDS became a formidable competitor technology for DSL and cable systems. However, the bandwidth of an MMDS cell is far from being adequate for all users in a 50 km radius. Thus, a new service type, called LMDS, is developed to work at higher frequencies [4]. Using 28-31 GHz in the U.S. and 40.5-42.5 GHz in Europe, LMDS is designed to provide high throughput. The cell size of an LMDS system is much smaller than its MMDS counterpart, ranging from 3 km to 5 km. Early LMDS cells support aggregate data rates of 34-38 Mbps per sector while later models increase this value to 36 Gbps ([4, 6]). LMDS systems are asymmetric and favor downlink over uplink. Their operational frequency range are more susceptible to problems like *Line of Sight* (*LOS*) connectivity requirement, rain and vegetation attenuation than MMDS systems.

The most important problem of LMDS systems are the lack of a standardization between systems of different companies, which causes considerable interoperability problems. To establish a standard for LMDS systems, IEEE formed The Work Group 16 which in turn developed the IEEE 802.16 standard in 2002 ([4, 7]).

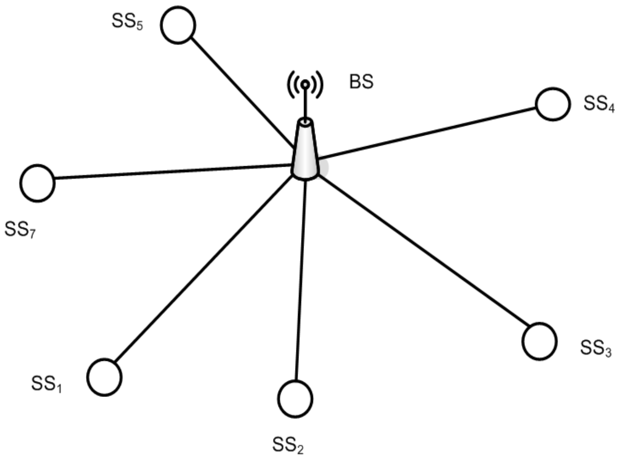


Figure 2.1. PMP topology of IEEE 802.16 [2]

The initial IEEE 802.16 standard provides connectivity for LOS subscribers in a PMP topology (Figure 2.1). Based on LMDS systems, the initial PHY layer of IEEE 802.16 works at the 10-66 GHz frequency band. The problems of LOS connectivity in urban settings forced the standard to develop another PHY layer for *Non-LOS* (*NLOS*) communications. This new PHY layer, developed as part of IEEE 802.16a, was introduced in 2003 [7]. In addition to the new PHY layer, IEEE 802.16a also introduced a second operation mode, the Mesh mode, to the standard. Since high multi-path propagation is required for NLOS communication and in the 10-66 GHz frequency band there is little multi-path propagation, a lower frequency band, 2-11 GHz, is chosen for NLOS operation [8]. Thus, IEEE 802.16a uses licenced and license-exempt frequencies in the 2-11 GHz band. After some amendments (mainly named under IEEE 802.16d) for both the standard and the PHY layer, IEEE 802.16-2004, was introduced in 2004 [2]. The recently finished standard, IEEE 802.16e, adds mobility support to the family [3]. The current version of the standard (IEEE 802.16-2005) includes both LOS and NLOS communication at the 10-66 GHz and the 2-11 GHz bands, respectively[1] .

IEEE 802.16 is a connection-oriented technology that includes protocols for both the PHY and *Data Link Control* (*DLC*) layer. Being a connection-oriented protocol, all transmissions in a IEEE 802.16 network are associated with connections. The

---

[1] We will refer to IEEE 802.16-2004 and IEEE 802.16-2005 as IEEE 802.16 as generally done in the literature.

Table 2.1. IEEE 802.16 PHY layer specifications [2]

Specification	Frequency Band	Optional Mechanisms	Description	Duplexing
WirelessMAN-SC	10-66 GHz		Main PHY specification of IEEE 802.16	TDD, FDD
WirelessMAN-SCa	<11 GHz	AAS, ARQ	Single Carrier specification for NLOS transmission	TDD, FDD
WirelessMAN-OFDM	<11 GHz	AAS, ARQ, Mesh	OFDM support for NLOS transmission, also used in Mesh topology	TDD, FDD
WirelessMAN-OFDMA	< 11 GHz	AAS, ARQ	OFDM support for NLOS transmissions	TDD, FDD
WirelessHUMAN	< 11 GHz	AAS, ARQ, Mesh, DFS	Main PHY specification for Mesh topology	TDD

connections are unidirectional and they can be unicast, multicast, or broadcast.

2.1. PHY Layer

Initially, IEEE 802.16 supported a single PHY layer, the WirelessMAN-SC PHY. Three additional PHY layers were developed for NLOS transmissions in the 2004 revision, one based on *Single Carrier* (*SC*) technology and two additional PHYs based on the *Orthogonal Frequency Division Multiplexing* (*OFDM*) technology (see Table 2.1). Channel bandwidth can be 20, 25, and 28 MHz for WirelessMAN SC (see Table 2.2). There are no fixed global channel bandwidth values for other PHY layers, but the available channel bandwidths are based on the frequency band that is used [2].

In LOS communication, aggregate raw data rate of the network is 36-135 Mbps

Table 2.2. 802.16 WirelessMAN-SC data rates (Mbps)[2]

Channel Bandwidth (MHz)	QPSK	16-QAM	64-QAM
20	32	64	96
25	40	80	120
28	44.8	89.6	134.4

based on the modulation and channel bandwidth used ([2, 9]). However, IEEE 802.16 provides up to 75 Mbps of aggregate raw data rate in NLOS communication ([2, 10, 11]). According to a performance analysis regarding the actual bandwidth of NLOS communication, IEEE 802.16 supports 10 Mbps for a 5 MHz-wide channel and 4.8-18.2 Mbps for a 6 MHz-wide channel [12]. In [13], Hoymann study the PHY and MAC layer throughput of an IEEE 802.16 network working at 5 GHz frequency band and using a channel bandwidth of 20 MHz. According to this work, the PHY layer gives a throughput ranging between 7 and 62 Mbps based on the modulation and coding scheme used. Also, it is found that MAC layer reduces the PHY layer throughput by 10%. The effects of optional MAC layer mechanisms, such as *Automatic Response Request* (*ARQ*) and packing, are also studied in this work. In [12], Ghosh *et. al.* propose various mechanisms to improve the current data rate, at least quadrupling the current data rate. The WiMAX forum on the other hand expects 15 Mbps maximum throughput per sector using 3.5 MHz channel bandwidth and 35 Mbps using 10 MHz channel bandwidth [14]. By using multiple adjacent channels, the bandwidth of the system can be improved up to 350 Mbps [10]. IEEE 802.16 networks can also be deployed using sectored antennas to further increase the overall bandwidth in a given area.

In order to ensure interoperability between WiMAX devices produced by different vendors, the WiMAX forum defined a profile for IEEE 802.16 devices. Two different frequency bands are used in this profile: 3.5 GHz and 5.8 GHz. The channel bandwidth is defined for these frequency bands as 3.5 or 7 MHz in 3.5 GHz and 10 MHz in 5.8 GHz. Among the PHY layers available, the profile uses WirelessMAN-OFDM with 256 carriers with either TDD or FDD [14].

2.1.1. PHY in PMP Mode

IEEE 802.16 supports both *Time Division Duplexing* (*TDD*) and *Frequency Division Duplexing* (*FDD*) to separate downlink and uplink communication. While BSs support full-duplex FDD, SSs may only support half-duplex FDD to minimize design costs. A continuous transmission of a IEEE 802.16 network is divided into fixed length parts called frames. In TDD mode, the frame consists of downlink and uplink subframes. In FDD operation mode, downlink and uplink subframes use different channels. The length of a frame is called the frame duration.

IEEE 802.16 includes several modulation schemes and *Forward Error Correction* (*FEC*) rates to cope with the variation in radio link quality due to weather, terrain, etc. The modulation techniques allowed in the standard varies with the PHY layer used. While *Quadrature Phase Shift Keying* (*QPSK*), 16-state *Quadrature Amplitude Modulation* (*QAM*) and 64-state QAM are supported in all PHY layers, a more robust modulation scheme, *Binary Phase Shift Keying* (*BPSK*), and a less robust one, 256-state QAM, are also supported in WirelessMAN-SCa PHY layer. FEC rates of 1/2 and 3/4 can be used for error correction. Together these values form a burst profile and each connection (either uplink or downlink) is described with a burst profile. Available burst profiles in the network are described with the *Uplink Interval Usage Code* (*UIUC*) for uplink and *Downlink Interval Usage Code* (*DIUC*) for downlink connections. Mapping of connections to these codes are broadcasted in *Downlink Channel Description* (*DCD*) and *Uplink Channel Description* (*UCD*) messages in each frame. In a single frame, an SS may have multiple connections with different burst profiles.

Connections are associated with burst profiles upon connection establishment. Transmissions start from the connection with the most robust burst profile and continue with decreasing robustness of the burst profiles in a frame ([2, 8]). In case, a link state changes an updated DCD or UCD message is sent by the BS in the next frame with new burst profiles for the connections. When the link state worsens, the connection switches to a more robust burst profile. On the other hand, if the link quality improves, the connection can switch to a less robust profile for higher bandwidth. While this change in burst profile is defined in the standard, it is not defined how the change will be handled. *Medium Access Control* (*MAC*) layer ETE delay can be used for burst profile changes. A comprehensive work in [15] shows that MAC layer ETE delay provides misleading information for handling the change in burst profile. However, network layer ETE delay can be used as a good metric for link adaptation purposes.

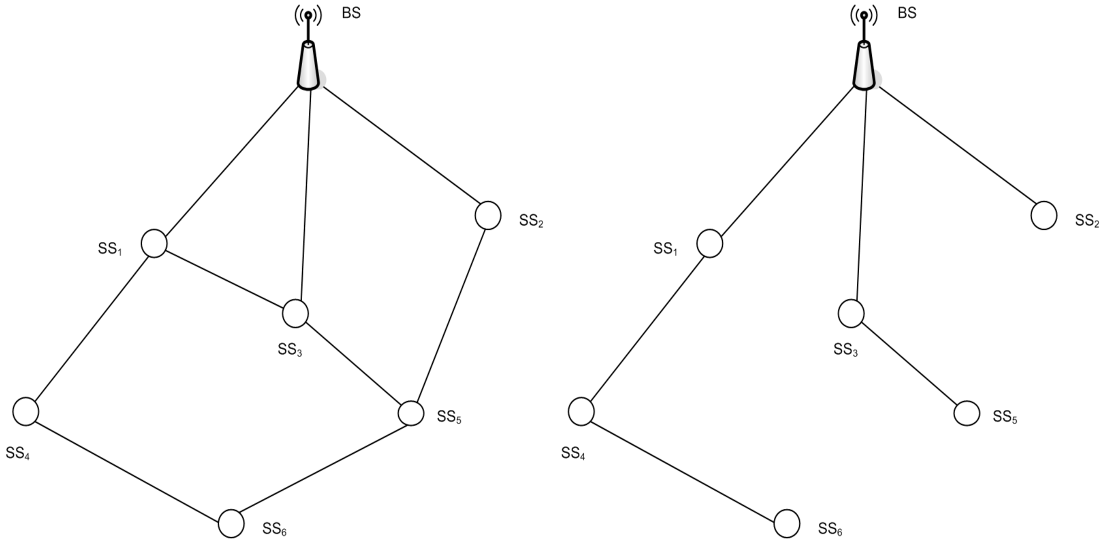
Broadcast and multicast connections in the uplink are essentially contention periods used for either bandwidth requests or initial ranging purposes. Each contenting SS randomly selects a *Transmission Opportunity* (*TO*) from the available TOs allocated to the connection in the uplink, and sends its request or message during the selected transmission opportunity. If more than one SS selects the same TO, a collision occurs and these colliding SSs retransmit their requests in the next frame until the transmission is successful or a timer expires. A more efficient ranging mechanism for *Orthogonal Frequency Division Multiple Access* (*OFDMA*) PHY layer is introduced in [16].

2.1.2. PHY in Mesh Mode

An SS is called *Mesh SS* (*MSS*) and the BS is called *Mesh BS* (*MBS*) in this mode. Unlike the PMP mode, transmissions are sent using the links between the nodes [2] . These links are bidirectional and are defined by 8-bit Link *Identifier* (*ID*)s. Each MSS has a hop count value that refers the minimum number of links a packet must travel to reach to the MBS. Upon initialization, a MSS establishes one link with each node in its range, these node are called the neighbour nodes of the MSS. Every MSS has a parent node in the Mesh mode. The neighbour nodes of the MSS with least hop counts among the neighbour nodes are called sponsor nodes. The parent node of a MSS is selected from these sponsor nodes. If the node is directly connected to the MBS, than the MBS is its parent node. The links between MSSs and their parent nodes are called form the scheduling tree (Figure 2.2). We call these links centralized links and all other links are distributed links. The performance of the scheduling tree greatly depends on the parent node selection in the initialization. IEEE 802.16 standard describes a method for selecting parent nodes. This method selects the node with the

---

[2] The node definition in WiMax Mesh mode refers to MSSs and the MBS in this mode.



(a) All links (b) Centralized Links

Figure 2.2. Mesh topology of IEEE 802.16

highest *Signal-to-Noise* (*SNR*) value among sponsor nodes as the parent node. This method does not guarantee that it finds the optimal scheduling tree. An analytical solution to find the optimal scheduling tree for the Mesh mode is described in [17].

Unlike the PMP mode, only TDD is supported in the Mesh mode.

2.1.3. Additional Mechanisms in the PHY Layer

IEEE 802.16 has an optional support for *Adaptive Antenna System* (*AAS*)s in the PMP mode. Using multiple antennas, BS can increase the signal range and quality. Whether there are non-AAS SSs in the network or not, AAS BSs have the ability to support non-AAS SSs. When there are both AAS and non-AAS SSs in a network, the downlink and uplink parts are divided into two parts for both types of SSs.

IEEE 802.16 also employs a *Dynamic Frequency Selection* (*DFS*) mechanism. With this mechanism, if there is a frequency conflict with another network, an IEEE 802.16 BS initiates a frequency change mechanism. BS and SSs actively sense the air medium for other data transmissions and available frequencies.

2.2. MAC Layer

MAC layer of the initial standard only supports PMP operating mode. The MAC layer of the current standard, the IEEE 802.16-2005 [3], also supports the Mesh operating mode. There are several differences between the two operating modes, mainly frame structure and bandwidth allocation methods.

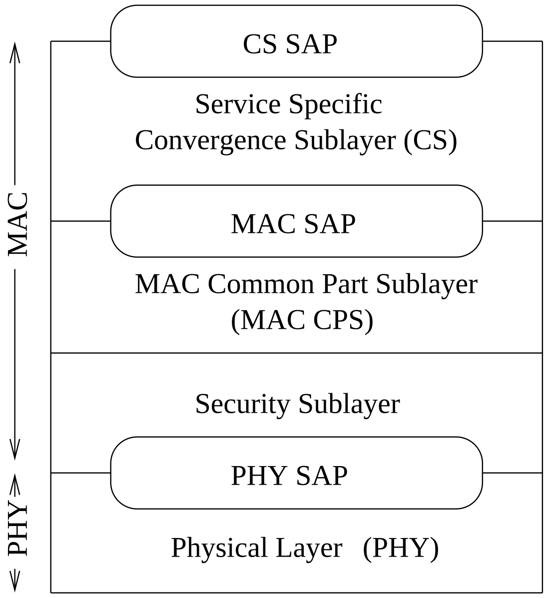


Figure 2.3. Layers of IEEE 802.16 [2]

Transmissions are described by *Service Flow* (*SF*)s and connections in IEEE 802.16. A transport service that provides the transmission of MAC *Packet Data Unit* (*PDU*)s between two nodes is called a SF. A unidirectional transmission between two nodes is defined by a connection in the IEEE 802.16 standard.

There are three sublayers in the MAC layer stack of IEEE 802.16 (Figure 2.3). Each sublayer has its own mechanisms and is responsible for different parts of the MAC layer.

2.2.1. Convergence Sublayer (CS)

CS is the highest sublayer of IEEE 802.16 MAC layer. Unlike the other two sublayers, there are different CSs in the MAC layer for different network layer protocols used in the network. Currently, IEEE 802.16 standard defines two CSs. ATM CS is used for handling ATM cells and Packet CS is used for Ethernet, PPP, and TCP/IP packets. In the transmitter side CS is mainly responsible for converting network layer packets into MAC *Segment Data Unit* (*SDU*) and vice versa in the receiver side. First, the CS maps the transmission parameters (including QoS parameters) into applicable IEEE 802.16 SFs. Then, repetitive information in the network layer header are eliminated by the *Packet Header Suppression* (*PHS*) mechanism. Lastly, the MAC SDU is sent to the lower sublayer.

2.2.1.1. ATM CS. There are two switching modes in an ATM network : VP - switched and VC - switched. In VP-switched mode, the connection is defined by *Virtual Path Identifier* (*VPI*). In VC-switched mode, the connection is defined by a VPI and *Virtual Circuit Identifier* (*VCI*) couple. The operation of the ATM CS differentiates between these two modes. It maps applicable identifiers based on the switching mode to SFs. These mappings are handled in connection establishment along with the QoS constraints of the VPI and VCI couple. After the connection establishment, ATM cells are mapped to IEEE 802.16 SFs based on their VPI or VPI and VCI values, depending on their switching mode. Since ATM cells with the same VPI or VPI and VCI values, have mostly the same information in their cell headers, these repetitive fields of the ATM cell header are suppressed if PHS is used in ATM CS. Only header information that is not dependent to VPI and VCI is not suppressed. Therefore, if the service provider wants to keep the intra-layer transparency, PHS mechanism should not be used.

2.2.1.2. Packet CS. The Packet CS defines a set of classifiers for network layer packets. These classifiers consist of protocol specific parameters (e.g. source and destination addresses, protocol version) that are used for mapping higher layer packets into IEEE

802.16 SFs. Priorities are assigned to classifiers. These priorities indicate in which order the classifiers are applied to packets. A SF can receive packets from more than one classifier. If a network layer packet does not match any of the classifiers in the Packet CS, that packet is discarded.

2.2.2. Common Part Sublayer (CPS)

The second part of the MAC layer is the CPS. This sublayer is the main sublayer of the MAC layer of IEEE 802.16. It handles access to the wireless medium, bandwidth allocations, connection establishment, connection management, and QoS management. MAC SDUs are fetched from CS sublayer via MAC SAP and converted into MAC PDUs. This sublayer is also responsible for packing and fragmentation of MAC SDUs. These mechanisms are explained below.

2.2.2.1. Packing & Fragmentation. The packing mechanism enables packing of multiple small MAC SDUs into one MAC PDU. MAC SDUs and PDUs can have either fixed or variable sizes. These size selections are based on the SF used. If the MAC SDUs have fixed sizes, the receiving node can easily calculate where the MAC SDUs begin and end. On the other hand, if MAC SDU size is variable a special subheader is used in each MAC PDU. This subheader, the packing subheader, is responsible for informing the receiver node how to access the MAC SDUs inside the MAC PDU.

In case of MAC SDUs that are larger than one MAC PDU, the transmitter node fragments the MAC SDU into multiple MAC PDUs. These MAC PDUs are sent in an orderly fashion. Similar to the packing subheader, a fragmentation subheader is inserted into each MAC PDU that belongs to the MAC SDU. This subheader informs the CPS of the receiving node in which order the MAC PDUs will be integrated. Packing and fragmentation can be used in conjunction with the ARQ mechanism. With the help of PHS and packing mechanisms, the standard eliminates bandwidth waste due to repeating information.

2.2.2.2. ARQ. In the PMP mode, during the establishment of SFs, the use of ARQ is negotiated between the SSs and the BS on SF basis. Whereas in the Mesh mode, ARQ is enabled on MAC PDU basis. While some MAC PDUs that are being transmitted on a link uses the ARQ mechanism, other MAC PDUs do not. SSs work in the PMP mode are not required to implement ARQ mechanism. However, SSs in the Mesh mode must implement ARQ mechanism whether they will use it or not.

2.2.3. Security Sublayer

The third sublayer, namely the Security Sublayer, is responsible for maintaining the security in the network. Security is maintained with encryption of data packets, secure key distribution via *Privacy Key Management* (*PKM*), authorization of PKM, and identification of nodes via X.509 profiles. Usage of these mechanisms are described with *Security Association* (*SA*)s. These SAs are identified by 16-bit *SA Identifier* (*SAID*)s. Each connection can be assigned a different SAID or a single SAID can be associated with a number of connections. Two types of SAs are defined: data SAs and authorization SAs. With the standardization of IEEE 802.16e [3], the security mechanisms are improved in order to cope with the threats arising from mobile profiles. *Johnston et al.* studied the security mechanisms in IEEE 802.16 in [18]. They have stated that the security mechanisms defined in the IEEE 802.16 standard have many flaws especially regarding authorization process since there is no explicit definition for authorization SAs in the standard. While the new security mechanisms introduced in IEEE 802.16e provides better protection against attacks, the authorization problem still exists and must be addressed.

2.3. Service Flows and Connections

As described above, a transport service that provides the transmission of MAC PDUs between two nodes is called a SF. Each SF has a 32-bit identifier called the SFID. An SS has a number of SAs at the same time, each with different service parameters. An SF defines various characteristics regarding the traffic supported by itself such as QoS parameters, the SA used in for the traffic, the choice of MAC SDU and MAC

PDU sizes and the choice of ARQ.

Three QoS parameter sets are associated with SFs; ProvisionedQoSParamSet, AdmittedQoSParamSet and ActiveQoSParamSet. These QoS parameter sets include the following QoS parameters: the maximum sustained traffic rate, the minimum reserved traffic rate, traffic priority, tolerated jitter, and the maximum latency parameters. The first parameter set is defined by higher level protocols (e.g. MPLS) and cannot be changed by the MAC layer of IEEE 802.16. If an SF has only ProvisionedQoSParamSet, it is called provisioned SF. The second parameter set is the AdmittedQoSParamSet. This parameter set is used by the BS to allocate resources for SFs. If a provisioned SF has its resources allocated, it becomes an Admitted SF. Lastly, if an SF is associated with an active connection (a connection that sends packets), it sends them based on the ActiveQoSParamSet of its SF. This kind of SFs are called Active SFs. Hence, both admitted and active SFs are associated with a connection.

Table 2.3. Management messages in PMP mode

Message Type	Management Connection Used
Ranging Messages	Basic Management
Burst Profile Change Messages	Basic Management
Reset Command	Basic Management
SS Basic Capability Messages	Basic Management
De/Re-register Command	Basic Management
ARQ Messages	Basic Management
Channel Measurement Reports	Basic Management
Adaptive Antenna System Messages	Basic Management
Uplink Channel Descriptor (UCD)	Broadcast
Downlink Channel Descriptor (DCD)	Broadcast
Downlink Access Definition (DL-MAP)	Broadcast
Uplink Access Definition (UL-MAP)	Broadcast
SS Network Clock Comparison	Broadcast
Fast Power Control	Broadcast
Privacy Key Management (PKM) Messages	Primary Management
Dynamic Service Addition / Alteration / Termination Messages	Primary Management
Multicast Assignment Messages	Primary Management
Simple Network Management Protocol (SNMP) Messages	Secondary Management
Dynamic Host Configuration Protocol (DHCP) Messages	Secondary Management
Trivial File Transfer Protocol (TFTP) Messages	Secondary Management

A unidirectional transmission between the BS and a SS is defined by a connection is the PMP mode of IEEE 802.16 standard. All admitted and active SFs are associated with a connection and every connection is associated with an SF. Connections are identified by 16-bit CIDs. Upon the initialization of a SS to the network, the SS setups its connections including three pairs of management connections in the PMP mode. These management connections are used for different management messages that are described in Table 2.3. Basic management connections are used for the most important and time-critical management messages. Primary management connections are used for more delay-tolerant management messages. Secondary management connections are only used in managed SSs. During the operation of the network, connections can be altered or terminated; as well as new connections can be established.

For link establishment in Mesh mode, the Link ID and four other link parameters are used to construct the CID. In this mode each MSS also has a 16-bit Node ID acquired from the MBS when the MSS is initialized. The Link ID and Node ID pair is used to generate requests and grants in the centralized scheduling of the Mesh mode.

2.4. Frame Structure

The transmission in an IEEE 802.16 network is divided into fixed length time units called frames. These frames have strict structures in both the PMP and the Mesh mode.

2.4.1. PMP Mode

Two different frame structures exist in PMP mode: FDD and TDD frame structure. In FDD frame structure, downlink and uplink subframes are transmitted in different frequencies (Figure 2.4). These subframes occupy the same frequency in the TDD frame structure but separated in time (Figure 2.5). Downlink and uplink subframes are used for transmissions originating from the BS and SSs respectively. Both subframes are formed of *Physical Slot* (*PS*)s. PSs allocated to the same connection are adjacent to each other and are called bursts.

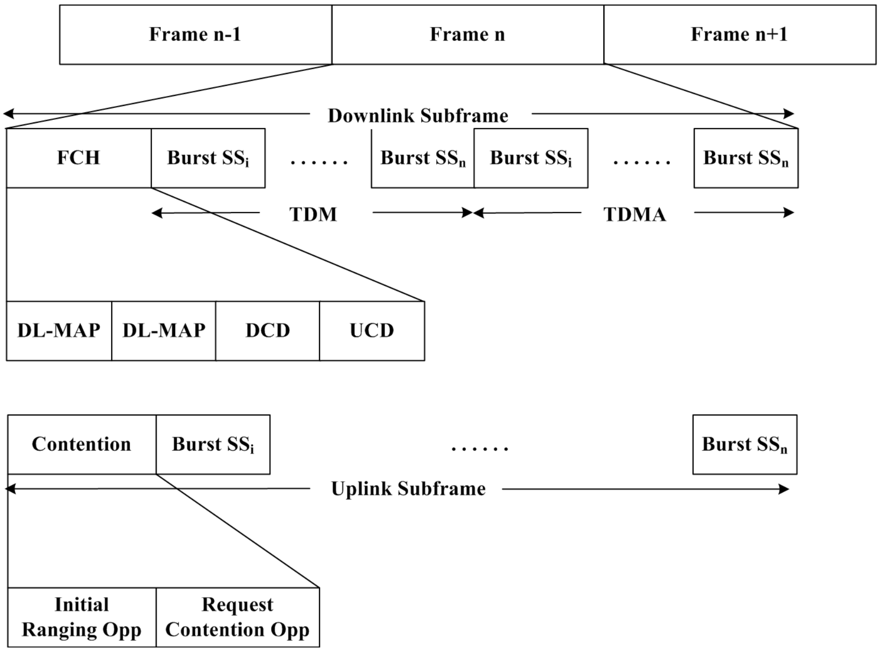


Figure 2.4. FDD Frame structure of the PMP mode of IEEE 802.16

Each downlink subframe includes a *Frame Control Header* (*FCH*) that is composed of the DL-MAP, UL-MAP, and the periodic DCD and UCD. *Downlink* (*DL*)-MAP and *Uplink* (*UL*)-MAP describe the mapping of the downlink and uplink PSs respectively to the bursts. DCD and UCD fields do not exist in every frame, instead they are transmitted periodically, once in a set number of frames. They describe the use of burst profiles to SSs. After the FCH, downlink data transmission is sent in bursts in a *Time Division Multiplexing* (*TDM*) manner. In the FDD mode, after the TDM downlink bursts, there is an additional *Time Division Multiple Access* (*TDMA*) part that is used for half-duplex SSs.

The data transmission is conducted in a similar manner in the uplink subframe as in the downlink subframe. PS allocations for broadcast connections form a contention-based part in the uplink subframe. The first contention-based part is used for the initial ranging and initialization process of the new SSs. Others are used by the SSs for their nrtPS and BE connections.

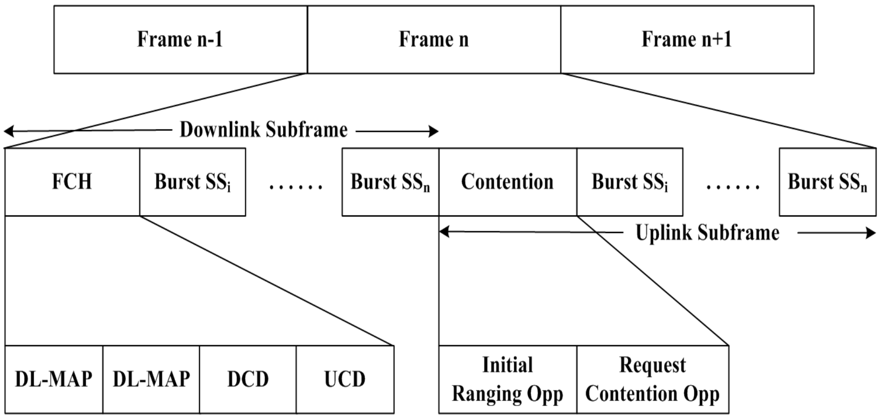


Figure 2.5. TDD Frame structure of the PMP mode of IEEE 802.16

The size of these contention periods affect the overall performance of the uplink subframe. Unnecessary large contention periods decrease the size of the uplink subframe usable by data transmissions. On the other hand, if the size of these contention periods are smaller than necessary, nrtPS and BE connections suffer unacceptably high delay values. In addition to these parts, there are periodic unicast PS allocations for each ertPS, rtPS, and nrtPS connection in the network. These allocations are small in size and are used to send only bandwidth requests to the BS. The effects of the contention window size are studied in [19].

2.4.2. Mesh Mode

The frame structure is quite different in the Mesh mode of IEEE 802.16 than the PMP mode. A frame can only be a TDD frame in the Mesh mode. In the Mesh mode, the frame is divided into two subframes; the control subframe and the data subframe (Figure 2.6).

The control subheader can be either a scheduling control subheader or a network control subheader. This subframe is composed of several TOs, that are in turn composed of seven OFDM symbols. Only one control message can be sent in one TO. Control messages of both scheduling methods are sent in the first type of control subframe. Some TOs are reserved for centralized scheduling control messages whereas the rest are used for coordinated distributed scheduling control messages. The second

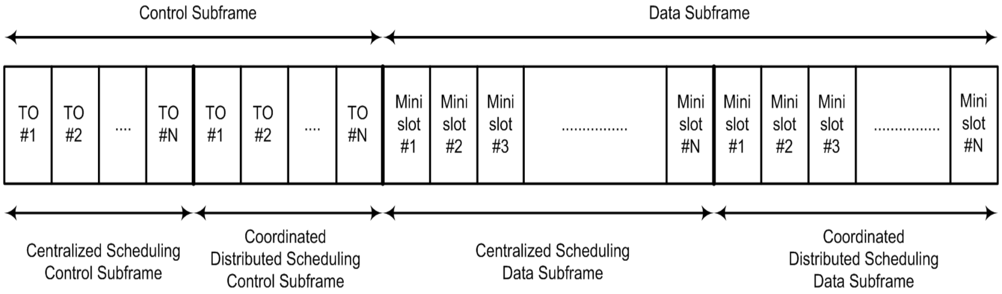


Figure 2.6. Frame structure of the Mesh mode of IEEE 802.16

subframe is used periodically similar to the UCD and DCD messages in the PMP mode (e.g. 1 network control subheader in 15 frames). The MBS informs MSSs about the data subframe channel usage by sending the burst profiles in the network configuration message. MSSs use entry messages to enter the network.

The data subframe is also divided into two parts among the scheduling methods. Both of these two data subframe parts are further divided into smaller minislots. The number of OFDM symbols defined by a minislot depends on the frame duration and the length of the control subframe. Allocations in the Data subframe are accomplished on minislot basis in both scheduling methods. The number of TOs in the control subframe, in other words the length of the control subframe, and the ratio of both subframe's allocations among the two scheduling methods are decided by the BS. These parameters can change during the operation of the network. In [20], *Schwingenschlögl et al.* has shown the effects of the control subframe length on the performance of the network.

While various modulation and coding rates can be used in the data subframe, only the QPSK modulation with 1/2 coding rate can be used in the control subframe.

2.5. Mobility

IEEE 802.16 is developed for fixed subscribers. Therefore it does not support mobile users. Leung *et al.* propose a mechanism that enables mobile users to IEEE 802.16 networks. A handover mechanism is developed for *Mobile Subscriber* (*MS*) based on the initialization procedure in the standard [21]. IEEE 802.16 Task Group E was established to standardize similar efforts and the standard is finished in December

2005. The standard allow MSs working in the 2-6 GHz frequency band with vehicular speeds up to 60 kmph and is expected to support data rates up to 30 Mbps [22].

Unlike a SS and a MSS, a MS can switch between BSs during a transmission with a handover mechanism. This handover procedure can be initiated by both the MS and its current BS. The handover procedure is similar to the initialization of a SS to the network. The MS may listen to the target BS before the handover and learn about its system parameters and thus speed up the handover procedure. Similarly, the MS may notify its target BS through the backhaul network. In this case, the target BS will allocate dedicated ranging opportunity to the MS that will be used by the MS instead of the contention-based ranging part. *Lee et al.* studied the handover process of the standard in [23], and propose a faster handover mechanism based on eliminating redundant work in the process.

Another major problem regarding the MSs is energy consumption. A mechanism called sleep mode is introduced in IEEE 802.16e to reduce the energy consumption of MSs. In this mechanism, the transmission between the BS and a MS is divided into two parts; interval of unavailability and interval of availability. During the first interval, the MS does not receive any transmission from the BS. BS buffers any packets arrived in this interval destined to the MS. During the second interval, the BS sends these buffered packets as well as the packets arrived in this interval to the MS. If there are no packets destined to the MS during an interval couple, the MS increases its sleep time and informs the BS about its new waking time. In [24], [25], and [26], it is shown that this power saving mechanism is effective.

A new working group, *Mobile Multihop Relay* (*MMR*) also known as IEEE 802.16j, is formed recently [27]. This standard will work in the PMP mode and allow SSs not directly connected to the BS to connect to the network. In order to achieve this goal there will be *Relay Station* (*RS*)s in the network. These RSs will be directly connected to the BS, and SSs will connect to the BS through these stations. RSs can only be able to relay a transmission. Data allocations in both downlink and uplink will be altered for to enable this transmission relaying.

3. QoS IN WiMAX

3.1. Traffic Scheduling in PMP Mode

Every SF is based on a scheduling service in the PMP mode of IEEE 802.16. These scheduling services define the nature of the data services supported, a rough QoS classification, and the set of allowed bandwidth request mechanisms for the connection. There are five different scheduling service classes available. Also, there are six QoS parameters defined in scheduling services. The applicability of these parameters vary between scheduling service classes (see Table 3.1).

UGS (*Unsolicited Grant Service*) This type of scheduling service supports real-time T1/E1 services and *Constant Bit Rate* (*CBR*) traffic. Upon connection establishment, the SS declares its bandwidth requirement to the BS for the connection. Then, the BS allocates exactly the requested amount of bandwidth to the connection in every frame. The bandwidth is always allocated to the SS regardless of the BS scheduler. The *Poll Me Bit* (*PMB*) in the grant subheader of UGS connections is used for non-UGS service requests. The bandwidth of the service is fixed and cannot be changed without restarting.

rtPS (*Real Time Polling Service*) While UGS supports real-time CBR traffic, rtPS supports real-time *Variable Bit Rate* (*VBR*) traffic. For each rtPS connection of an SS, the BS assigns a periodic request opportunity in the uplink subframe. Thus, the connection never contends for bandwidth allocation. The size of the requested bandwidth varies from time to time, up to a limit set during the setup of the connection. Due to this request/grant mechanism, there are some overhead packets for a rtPS connection.

nrtPS (*non-Real Time Polling Service*) nrtPS connections carry non-real-time traffic. The same polling mechanism used for rtPS connections is also used for nrtPS. Unlike rtPS, the connection may also enter contention for non-periodical bandwidth allocation request. Since these connections are not as important as rtPS connections and they have the ability to enter contention for bandwidth alloca-

Table 3.1. Scheduling services of IEEE 802.16

	UGS	rtPS	nrtPS	BE	ertPS
Preferred Traffic Type	CBR	VBR	VBR	ABR	VoIP
Periodic Polling Allowed	-	+	+	-	+
Usage of PMB Allowed	-	+	+	+	+
Usage of Contention Periods Allowed	-	-	+	+	-
Max. Sustained Traffic Rate	+	+	+	+	+
Max. Latency	+	+	-	-	+
Tolerated Jitter	+	-	-	-	-
Request / Transmission Policy	+	+	+	+	+
Min. Reserved Traffic Rate	+/-	+	+	-	+
Traffic Priority	-	-	+	+	-

tion requests, the polling periods of nrtPS connections are longer than that of rtPS connections.

BE (*Best Effort*) This type of service can send bandwidth allocation requests only using contention. BS never allocates dedicated request opportunities to the SS for BE connections.

ertPS (*Extended Real Time Polling Service*) In [28], it is shown that current scheduling services are not appropriate for services like *Voice over IP* (*VoIP*). Addressing this issue, the latest standard of IEEE 802.16 introduced ertPS scheduling service. ertPS is similar to UGS since it does not have any bandwidth request mechanism and in every frame the BS allocates bandwidth for the connection. However, the bandwidth allocated to the connection can change in time, similar to rtPS. An ertPS connection can decrease or increase its allocated bandwidth based on the traffic.

The performance of these scheduling services is evaluated in [29]. In this work, it is shown that average uplink delay is greater than downlink delay because of the polling and request mechanisms. Also, the requirements of these scheduling service classes are satisfied with the current request and grant mechanisms stated in the standard. Application layer services use the most appropriate of these five scheduling service types for the given service.

3.2. Traffic Scheduling in Mesh Mode

Two scheduling methods are defined in the Mesh mode of IEEE 802.16: centralized scheduling and distributed scheduling. While centralized scheduling can be used alone, distributed scheduling can only be used with centralized scheduling. Centralized scheduling is similar to the PMP mode. Each SS sends its bandwidth request to the BS and all the scheduling in the network is managed by the BS. A node that is not directly connected to the BS send its bandwidth request message to its parent node who forwards it over its own parent node towards BS. Each SS requests bandwidth only for the links on the scheduling tree and only for uplink in these links. This mode is generally used for Internet traffic in the network.

Distributed scheduling entails two methods: *Coordinated Distributed Scheduling* (*CDS*) and *Uncoordinated Distributed Scheduling* (*UDS*). As opposed to centralized scheduling, none of these methods has a single point of scheduling control. Instead, every SS distributes the scheduling information of its 1-hop neighbors and its own scheduling information to its 1-hop neighbors. Thus, each node knows the scheduling scheme in its 2-hop neighborhood and makes its scheduling accordingly. Both of the methods use a three-way handshake mechanism for bandwidth allocation. The main difference between these two methods is that the scheduling information is sent in a collision free manner in CDS whereas in UDS collisions of scheduling messages are possible. Distributed scheduling is generally used for intranet traffic in the network.

3.2.1. Centralized Scheduling

In the CS, CSCH messages are used for both requests and grants. Also, the BS informs SSs about the topology of the network using periodic CSCF messages. SSs that do not have any child nodes, send their uplink bandwidth requests to their parent

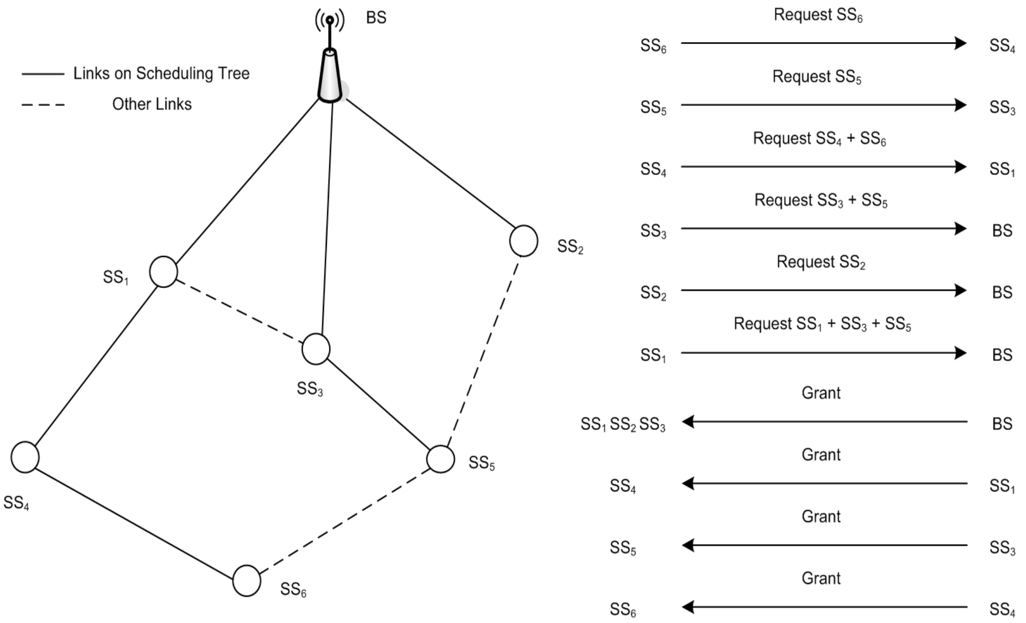


Figure 3.1. Example of centralized scheduling

nodes via CSCH messages. SS with child nodes take the uplink bandwidth requests from all their child nodes and generate a CSCH message including these requests as well as their own uplink request. In a network consisting of N SSs, since each CSCH message uses one TO, N number of TOs are needed in order for all SSs to send their bandwidth requests to the BS. The control subframe has a maximum size in the standard ($Csto_{max}$), also some of these TOs should be reserved for the CDS. In medium to large topologies where N value is greater than the available number of TOs in the CS part of the control subframe ($Csto$), requests are sent in more than one frame. The number of frames required for every SS to send its uplink bandwidth requests is denoted by $ReqFrCnt$ and can be calculated by Eq 3.1.

$$ReqFrCnt = \left\lceil \frac{N}{Csto} \right\rceil \tag{3.1}$$

The BS generates a grant using the uplink bandwidth requests from SSs and the downlink bandwidth requests from itself via a mechanism called the BS scheduler. The grant is then distributed to the SSs again by using CSCH messages. Upon receiving the grant message, SSs that have child nodes forward this grant to their child nodes. This forwarding continues until every node acquires the new grant. Similar to the request sending procedure, the grant messages may exceed one frame. The number of frames required for the grant messages to reach every SS in the topology is denoted by $GrtFrCnt$ and can be calculated by Eq 3.2 where $NodeWChldCnt$ denotes the number of SSs with child nodes. Thus, each request - grant period takes $ReqFrCnt+GrtFrCnt$ number of frames.

$$GrtFrCnt = \left\lceil \frac{NodeWChldCnt + 1}{Csto} \right\rceil \tag{3.2}$$

When sending uplink requests, the SS with the highest Node ID sends its CSCH message first and the SS with the lowest Node ID sends last. Grant messages propagate in the network in the reverse order of the request messages. Starting from the BS, continued by the SS with the lowest Node ID that has at least one child and ends with the SS with the highest Node ID with at least one child (Figure 3.1). Using the topology information gathered by the CSCF messages together with this request and grant sending sequences, the SSs send their CSCH messages without any collisions.

In order to decrease the packet size of the CSCH messages, only a 4-bit field is allocated for each uplink flow request rup_i and grant gup_i, in request and grant messages respectively. There is an additional 4-bit field called Flow Exponent (f) in every CSCH message. The bit values of requests and grants can be calculated using these fields by Eq 3.3.

$$\begin{aligned} ReqUp_i &= rup_i \cdot 2^{f+14} \\ GrtUp_i &= gup_i \cdot 2^{f+14} \end{aligned} \tag{3.3}$$

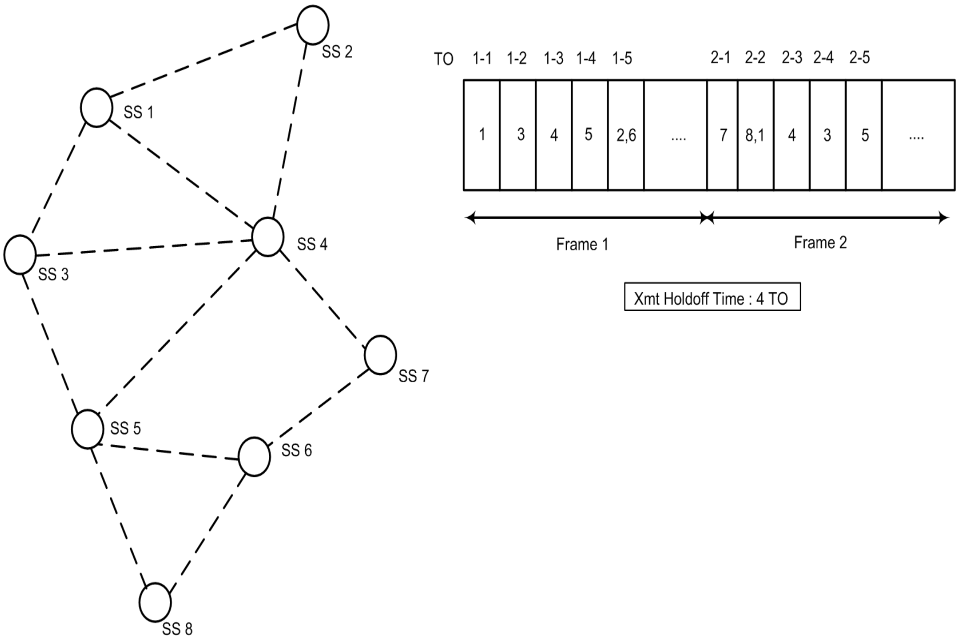


Figure 3.2. Example of DSCH TO timing mechanism

3.2.2. Coordinated Distributed Scheduling

Scheduling information is sent using the DSCH messages in a non-collision manner in this scheduling method. Since a single point of control does not exist in CDS, each SS has to calculate a unique TO to send its DSCH message to avoid collisions. All SSs in the network know the complete topology of the network from the periodic CSCF messages. Node ID's of the SSs are also included in this message. In order to choose a unique TO, each SS calculates a hash value for each of its 2-hop neighbours including itself using SS's Node IDs and the number of the TO in question (starting from the first applicable TO for itself). For each TO, the SS whose hash value is the greatest is assumed to have won the contention and uses this TO to send its DSCH message. Each SS continues to calculate these hash values until it wins the contention for one TO. In order to be fair among SSs this mechanism introduces an additional backoff mechanism. After winning for a TO, an SS cannot enter contention for a given number of TOs (Xmt Holdoff Time). However, this mechanism requires the SSs to know when each of their 2-hop neighbours will send their next DSCH messages. In the standard, this issue is mentioned and several fields are assigned in the DSCH packet format for this purpose

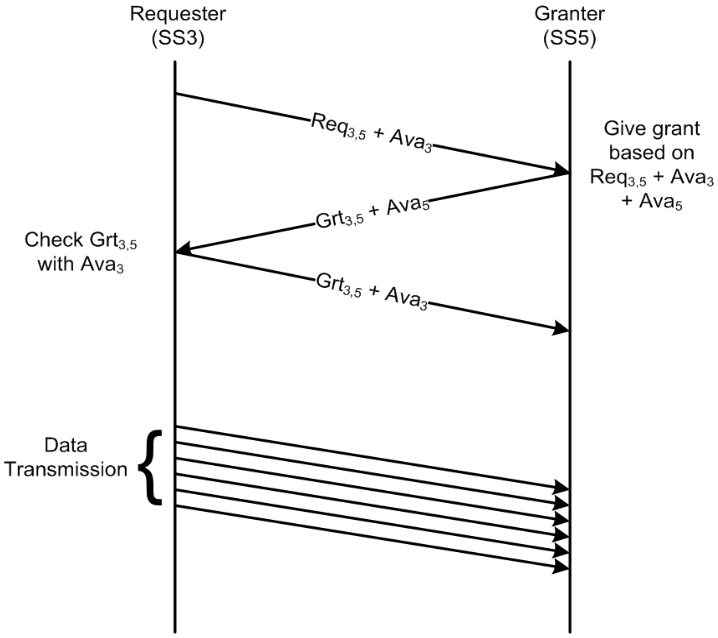


Figure 3.3. Example of coordinated distributed scheduling

but a detailed mechanism regarding the usage of these fields is not included. In this work, we assume that each SS knows each of its 2-hop neighbours' next DSCH sending TO using the fields in the DSCH message reserved for this purpose (Figure 3.2).

Each DSCH message includes requests, availabilities and grants. Every SS sends requests (if necessary) for every distributed link it has in the network when its DSCH message sending TO comes. Upon receiving a DSCH message, a SS checks if there is a request for the link between itself and the transmitting SS. Then, the receiving SS checks whether it can grant the request in several steps. Each SS knows in which minislots of the DS part of the data subframe it is available to transmit and/or receive data. When sending a DSCH message, a SS includes its availabilities into the message. Thus, the granter knows the availabilities of both the requester and itself. Based on these information, the receiving SS generates a grant to the request if a grant can be given to this request. This grant is sent to the requester with a DSCH message. Since the availability list of the requester sent in the first DSCH message might be stale, the requester checks the grant within its current availability list if it is still available for this grant. If the requester is still available for this grant, it sends the grant again to the granter to acknowledge the grant. Thus, the three-way handshake is complete and the data transmission commences as agreed between the two SSs (Fig 3.3). A recent work, [9] shows the effects of different variables in the performance analysis of distributed scheduling Mesh mode.

3.2.3. Uncoordinated Distributed Scheduling

This scheduling method is very similar to CDS in many aspects. The same three way handshake and DSCH message format used in CDS is also used for data allocation in UDS. However, the DSCH messages are not sent in the control subframe in UDS. SSs use minislots of distributed scheduling part of the data subframe according to their availability lists to send their DSCH messages. Unlike the CDS, there is no guarantee that the DSCH messages will not collide. If there does not exist any available minislots in the distributed scheduling part of the data subframe for a given SS at a frame, UDS cannot be used.

3.3. BS and SS Schedulers in The Literature

QoS schedulers in both the BS and SS sides are left unstandardized in the original standard. These schedulers have a significant effect on the overall performance. The BS allocates bandwidth on SS basis rather than per connection in the PMP mode. Thus, it does not specify for which connection the allocated bandwidth will be used. The SS decides the order in which the connections send their data. This distributed structure handles fairness between SSs, which in turn improves overall performance.

In the literature, there are several proposals for BS and SS schedulers. In [30], a SS scheduler is presented in which connections with the same scheduling services are integrated and different queuing policies are applied to the queue of each scheduling service. The authors propose using *Wireless Packet Scheduling* (*WPS*) for rtPS connections, *Weighted Round Robin* (*WRR*) for nrtPS connections and *First In First Out* (*FIFO*) scheduler for BE connections. Wongthavarawat *et al.* propose a BS scheduler in [31]. In this proposal, arrival times of rtPS PDUs are sent to the BS through the UGS connection of the same SS. Also, the BS scheduler applies different queuing policies to different scheduling services; EDF scheduling for rtPS connections and WFQ scheduling for nrtPS connections. In [32], Jiang *et al.* develop another BS scheduler using token buckets to characterize traffic flows. In [29], a WRR scheduler is used for uplink bandwidth allocation in the BS scheduler and a DRR scheduler is used in SS schedulers. The DRR scheduler is also used for the downlink bandwidth allocation in the BS scheduler. A queue state aware SS scheduler for polling service connections is proposed and its performance is analyzed in [33]. This scheduler informs the packet source of its queue status and tries to control the packet arrival rate.

In [34], a BS scheduler for the Mesh mode is introduced. This scheduler introduces a node ordering mechanism among the nodes with same hop count from the MBS. Moreover, an *Space Division Multiple Access* (*SDMA*) mechanism is used to further increase the throughput in the network. Since this BS scheduler uses SDMA, several changes should be done to the standard for the BS scheduler to work.

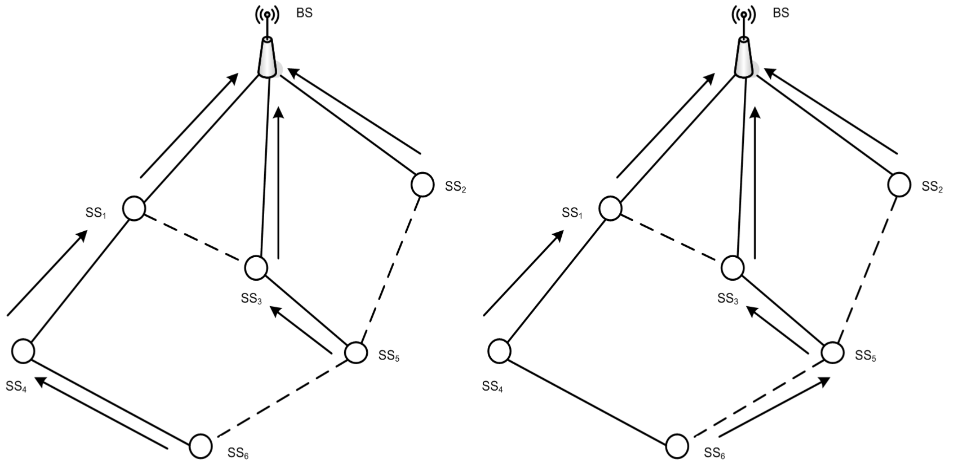
4. QUEUE AWARE ROUTING

The Mesh mode of WiMAX carries both Internet and intranet traffic. While providing Internet access is the main goal, considerable intranet traffic is also expected. As stated above, the size of the data subframe that is reserved for CS and DS is determined by the BS. In order to allow intranet traffic, the BS should reserve some of the data subframe to DS. However, the BS has no information on the utilization of the DS part of the data subframe. We propose selecting the size of the data subframe allocated to DS, by using a number of minislots large enough for average intranet traffic.

Being a broadband wireless access network, the demand for Internet is expected to exceed that of intranet traffic in a WiMAX network. Our method basically allows the DS part of the data subframe for Internet traffic usage when the CS part of the subframe suffers from congestion. Since there is no BS scheduler specified in the standard, we develop a BS scheduler for the Mesh mode. Also, there does not exist any requester for distributed scheduling in the standard. Our work also includes a requester for the distributed scheduling of the Mesh mode.

4.1. Queue Aware Routing

In QAR, if a SS has more than one potential parent nodes, a second node among the potential parent nodes is selected as the pseudo parent after selecting the parent node. In the case of congestion in the CS, the SS changes the routing of the Internet traffic from its parent node to its pseudo parent node. Since the link between the SS and its pseudo parent node is a distributed link, the centralized link usage for this node's Internet traffic decreases (Figure 4.1). Thus, the total traffic introduced to the CS decreases, and the system goes back to an uncongested state. In order not to hinder the intranet traffic unnecessarily, the SS switches back its routing of Internet traffic from its pseudo parent to its parent node when the congestion in the CS ends.



(a) Uncongested (b) CS congested SS6 using pseudo parent

Figure 4.1. Internet traffic routing

We develop two approaches to determine when to change the route from the parent node to the pseudo parent node, and when to revert back to the parent node from pseudo parent node. We call these two approaches *Centralized Queue Aware Routing* (*CQAR*) and *Centralized, Distributed Queue Aware Routing* (*CDQAR*). While determining the start of a congestion, both approaches use the queue length of the centralized link. When the queue length of this link exceeds a certain threshold queue length (Centralized Link Congestion Threshold, $CLCth$), the SS starts to route all of its Internet traffic from its pseudo parent. The CDQAR also checks the length of the distributed links, if the queue length of any of these links exceeds a threshold (Distributed Link Congested Threshold, $DLCth$), then it means at least one distributed link is in use and the SS cannot use its pseudo parent for Internet traffic.

In CQAR, the end of the congestion is identified by checking the queue length of the centralized link again. We use a second threshold value (Centralized Link Uncongestion Threshold, $CLUth$) for switching the Internet traffic routing back to the parent node to avoid instability in the system. On the other hand, CDQAR uses the pseudo parent unless there is a congestion in one of its distributed links. When the queue length of any one of the SSs distributed links exceeds $DLCth$ the SS switches back to using its centralized link for its Internet traffic regardless of the status of its centralized link.

Using these parameters as decision parameters, CQAR is only expected to handle short period congestions, while CDQAR can handle both short term and long term congestions. CDQAR is also aware of the status of the distributed links. If any of the node's distributed links is congested, CDQAR changes the route back to original. However, this approach can only handle distributed traffic initiated from the node itself. If a 1-hop neighbour node is trying to send a packet to the node that uses its pseudo parent, CDQAR cannot sense the congestion in the distributed links. Also, in the case of traffic between a 1-hop neighbour and a one or two hop neighbour, if there is a congestion in this traffic this congestion cannot be sensed in CDQAR. Since CQAR uses the pseudo parent briefly, CQAR is better at handling these kind of problems.

In [35], Cheng et al. propose a method called CDS in which the data subframe is not divided into two parts as in the standard. Thus, the Internet and intranet traffic can use the data subframe completely. However, their work has two significant shortcomings. First, they assume that the BS scheduler can give grants using individual minislot basis. Since the standard uses reduced 4-bit values for grants minislots can only be allocated consecutively to SSs. Thus, the CSCH packet format must be changed for their method to work. Furthermore, since there is no fixed minislot size for distributed scheduling the number of minislots available for distributed scheduling may vary greatly from frame to frame. Thus, for high SR values only low Per values can be used, which will results in high delays for intranet traffic.

Our method only changes the routing of the data packets in the network. We do not propose any changes to the MAC layer mechanisms of IEEE 802.16 standard. Thus, our method can be implemented in current and future SS and BS devices easily since both QAR approaches abide by the standards.

4.2. Hop Count Aware BS Scheduler

To handle Internet traffic scheduling, we develop a *Hop Count Aware* (*HCA*) BS Scheduler. Since the CS is mainly used for Internet traffic the requested traffic is destined to the BS. If we denote the hop count of SS_i as Hc_i, the uplink bandwidth request (in bits) of this SS with $ReqUp_i$, and the uplink bandwidth grant (in bits) for this SS with $GrtUp_i$, this grant actually consume, $GrtUp_i \cdot Hc_i$ bits from the network's total bandwidth. Since the BS knows the Hc_i values for each SS in the network, it can calculate how much traffic is actually necessary for each SS ($ReqUp_i \cdot Hc_i$), and use these values instead of the $ReqUp_i$values. The HCA uses this approach in both uplink and downlink grants.

HCA also prioritizes SSs based on their hop counts, Hc_i values, starting from SSs with $Hc_i = 1$, and lastly SSs with $Hc_i = Max_{Hc}$ where Max_{Hc} denotes the highest hop count in the network. Thus, the HCA tries to increase the utilization of the network. The total amount of actual bandwidth usage of the SSs with the current hop count is compared with the remaining available minislots in the network. If the total amount is less than or equal to the remaining available minislots, each request is granted completely. Otherwise, the SSs with the current hop count are listed randomly using a uniform distribution, and their requests are checked in the order in this list. In order not to differentiate between downlink and uplink requests, each SS appears twice in this list, the first is for downlink and the second is for the uplink.

Using HCA, when the total amount of request exceeds the system capacity, the bandwidth requests of SSs with $Hc_i = Max_{Hc}$ are not fully granted. With increasing total requests, the bandwidth requests of SSs with Hc_i values lower than Max_{Hc} are not granted with decreasing Hc_i values.

4.3. Distributed Scheduling Requester

While the DSCH packet format and the three-way handshake is defined in the standard, neither a requester nor a granter is defined. We develop a requester that handles the data traffic with low delay and jitter values while reducing the number of requests. A request in a DSCH message includes two parameters: SR and Per. The SR parameter defines how many minislots this transmission consists of in each frame. The Per parameter defines how many frames this transmission lasts.

Firstly, our requester records the amount of traffic received from the higher layers during each frame in minislots. When a SS is sending its DSCH message, a SR and Per value is calculated for each distributed link the SS has, based on the recorded received traffic values. The variance of the last M values of this received traffic are calculated. If the variance is below a certain threshold, the requester sets the Per value equal to M and the SR value to the average of these M values. If the variance is below the threshold, another variance is calculated using the last $\frac{M}{2}$ values and compared with another threshold. This calculation continues unless the variance is below its related threshold, or $\frac{M}{2}$ is equal to 1. After a request is granted, the requester will not try to send any requests for this link until the current request's Per value is expired. Thus, if the traffic is bursty, our requester sends many requests with different number of SRs in each request with small Per values. On the other hand, if the traffic does not change for many frames, a few requests are sent using SR values close to each other on each request with high Per values.

We use a simple granter in the CDS. The granter checks for a vacancy for the request in the two availability lists. It gives a grant if it finds an appropriate vacancy in the next 256 frames. Otherwise, the granter does not give a grant to the request, implicitly meaning it cannot give a grant for such a request.

5. SIMULATION RESULTS

5.1. Simulation Scenario and Parameters

We test our proposed solution in a moderately dense topology consisting of one BS and 10 SSs (Figure 5.1). The maximum hop count in the topology is three, and the distribution of number of SSs based on the hop counts are three SSs at 1-hop distance, four SSs at 2-hop distance, and three SSs at 3-hop distance.

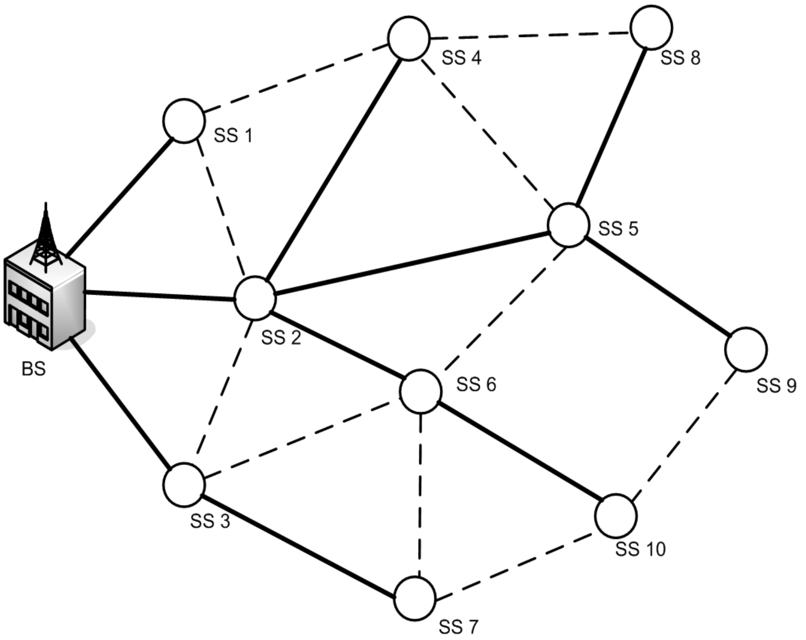


Figure 5.1. Simulation topology

We assume free space propagation model for the PHY layer model. Thus, we use the most robust PHY layer available in the standard, 64-QAM with 3/4 coding rate, for the data subframe. The frame duration is chosen as 5 msec in our simulation. To reduce the $ReqFrCnt$ and $GrtFrCnt$ values, we set the $Csto$ to its maximum available value in the standard. The rest of the simulation parameters are listed in Table 5.1.

There are three traffic types in the simulation; downlink Internet, uplink Internet, and intranet traffic. All traffic follows Poisson distribution. The downlink Internet traffic is considered as background traffic and exists during the whole simulation. The uplink Internet and the intranet traffics are implemented as ON/OFF traffics

Table 5.1. Simulation parameters

Frame Duration (msec)	5
Minislot Length (OFDM Symbol)	2
Data Subframe Modulation	64-QAM
Data Subframe FEC Rate	3/4
SS Count	10
Queue Size (Centralized Link) (Kbit)	1000
Queue Size (Distributed Link) (Kbit)	250
Max_{Hc}	3
$Csto$ (TO)	15
$Csto_{CS}$(TO)	10
$Csto_{DS}$(TO)	5
$ReqFrCnt$ (Frame)	1
$GrtFrCnt$ (Frame)	1
Minislot Count in Data Subframe	155
Minislot Count Reserved for CS	125
Minislot Count Reserved for DS	30
CS Capacity	44 Mbit/sec
DS Capacity	11 Mbit/sec
$CLCth$	30%
$CLUth$	10%
$DLCth$	40%

with changing starting times and durations among different SSs. Since the distributed scheduling utilizes SDMA, we only introduce intranet traffic to and from the 3-hop SSs. The simulation duration is 3 minutes, which results in 36000 frames.

SSs record their centralized link queue lengths once every second in our mechanism. Congestion control checking mechanism runs once in every 5 seconds using the queue lengths recorded in the last 5 seconds.

Since in a realistic system SSs do not enter the network at the same time, the

Table 5.2. Traffic parameters

Downlink Traffic Rate per SS (Mbit/sec)	1.0
Uplink Traffic Rate per SS (Mbit/sec)	1.1
Intranet Traffic Rate (SS_{10} -> SS_9) (Mbit/sec)	8.8
Intranet Traffic Rate (SS_7 -> SS_{10}) (Mbit/sec)	1.0

periodic centralized link queue length checking should be asynchronous. In our simulation SS_i, start using the QAR system at the $QARch_i^{th}$ second. The $QARch_i$ is calculated for each SS using Eq 5.1.

$$QARch_i = i \ mod \ 5 \tag{5.1}$$

Table 5.3. Traffic timing parameters

	Start Time (sec)	ON Duration (sec)	OFF Duration (sec)
Up SS_1	4	41	20
Up SS_2	3	46	20
Up SS_3	9	44	20
Up SS_4	1	48	20
Up SS_5	6	45	20
Up SS_6	8	49	20
Up SS_7	2	43	20
Up SS_8	10	50	20
Up SS_9	7	47	20
Up SS_{10}	5	42	20
Intranet SS_{10} -> SS_9	25	30	30
Intranet SS_7 -> SS_{10}	25	30	30

Table 5.4. Alternative Routing Enabling and Disabling Times (CQAR)

	SS8	SS10
ON	18	15
OFF	23	20
ON	28	25
OFF	33	30
ON	38	35
OFF	43	40
ON	-	45
OFF	-	50
ON	88	90
OFF	93	95
ON	98	100
OFF	103	105
ON	108	110
OFF	113	115
ON	158	155
OFF	163	160
ON	168	165
OFF	173	170

5.2. Simulation Results

The simulation starts with only downlink traffic. After the first second SSs starts generating uplink traffic and after the 10th second all the SSs generate uplink traffic. Thus, the network becomes congested. This congested state lasts until two SS stops its uplink traffic at 45th second. This congested Internet traffic state occurs twice again in the simulation between 77th- 106th seconds and 150th - 167th seconds. Both of our methods change the routing of the Internet traffic in the congestion cases. Only the three-hop SSs with pseudo-parent change their route since only they are affected from congestion. The routing changes are summarized in Table 5.4 and 5.5.

Table 5.5. Alternative Routing Enabling and Disabling Times (CDQAR)

	SS_8	SS_{10}
ON	18	15
OFF	-	30
ON	-	35
OFF	-	40
ON	-	45
OFF	-	50

In order to show the effects of our mechanism in a worst case scenario, we introduce loaded intranet traffic when the Internet traffic is already congested at 25th second.

The load of the system on 1-hop SSs is minor. As can be seen in Figure 5.2, the average downlink ETE delay of 1-hop SSs does not change between the three methods. Only in the CDQAR there is a very slight improvement of 0.5 msecs with respect to the other two methods.

The effects of both QAR mechanisms can be seen in average downlink ETE delay of 2-hop SSs in Figure 5.3. However, since the load does not affect 2-hop SSs as in the caseof 1-hop SSs the improvement of both our methods is still marginal. The CQAR reduces the downlink ETE delays of 2-hop SSs by 1-1.5 msecs and the CDQAR reduces the ETE delays by 2.5-3 msecs.

In case of congestion in the centralized links, firstly the three-hop SSs suffer high ETE delays and packet drops. In Figures 5.4, 5.5, and 5.6, during the three congestion periods the ETE delay of all three three-hop SSs increase dramatically. CQAR decreases the high ETE delay values but its reduction is not stable since after changing the Internet traffic routes SSs 8 and 10 only use the alternative routes until the packets at their centralized link queues are sent and their centralized queue lengths drop below 10 % of their capacity. Then, they switch back their original routes. However, in CDQAR, after the first time the centralized link queues of SSs 8 and 10's exceed

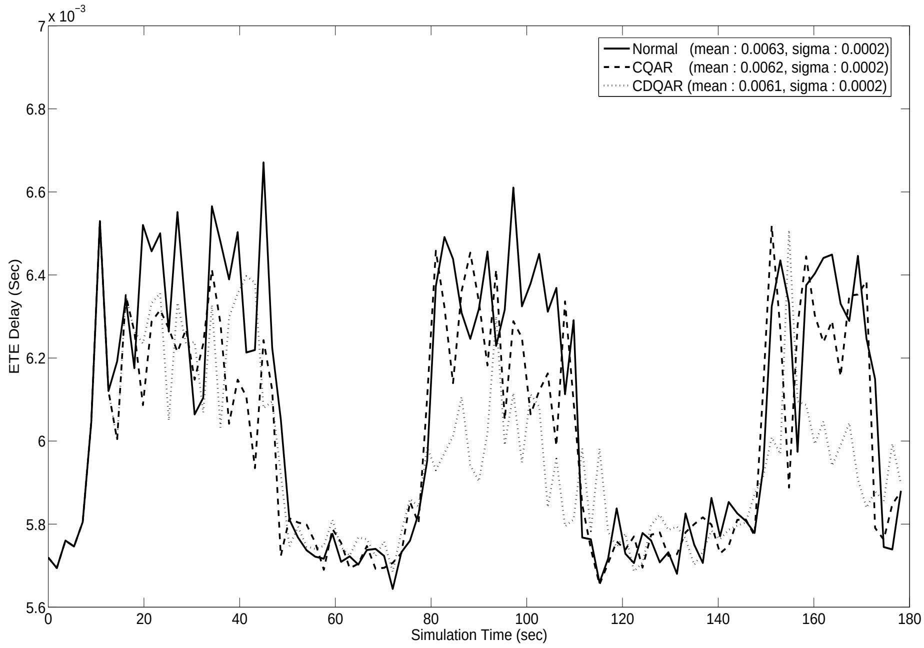


Figure 5.2. Average ETE delay of 1-hop SSs (downlink internet traffic)

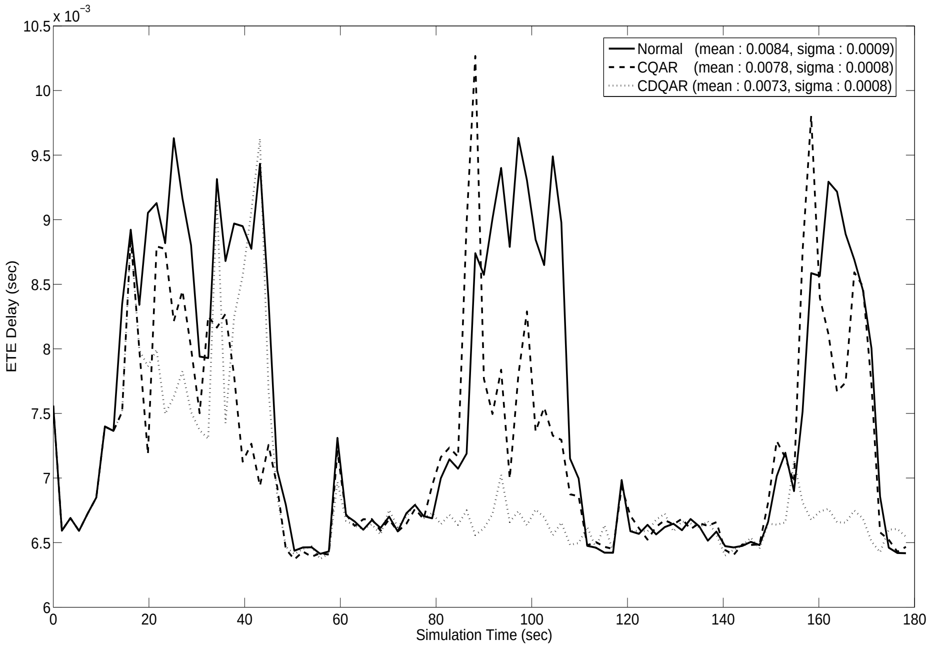


Figure 5.3. Average ETE delay of 2-hop SSs (downlink internet traffic)

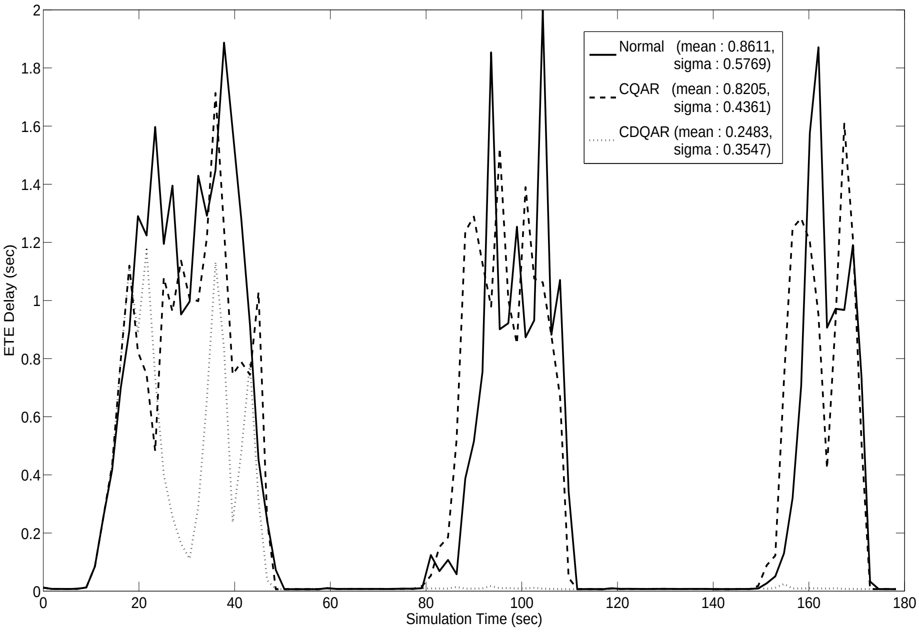


Figure 5.4. ETE delay of SS_8 (downlink internet traffic)

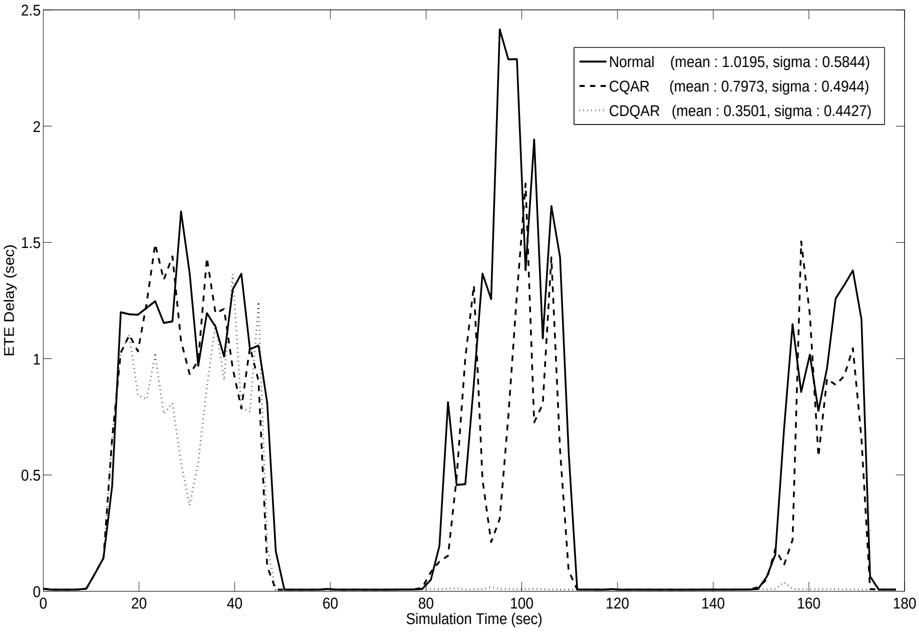


Figure 5.5. ETE delay of SS_9 (downlink internet traffic)

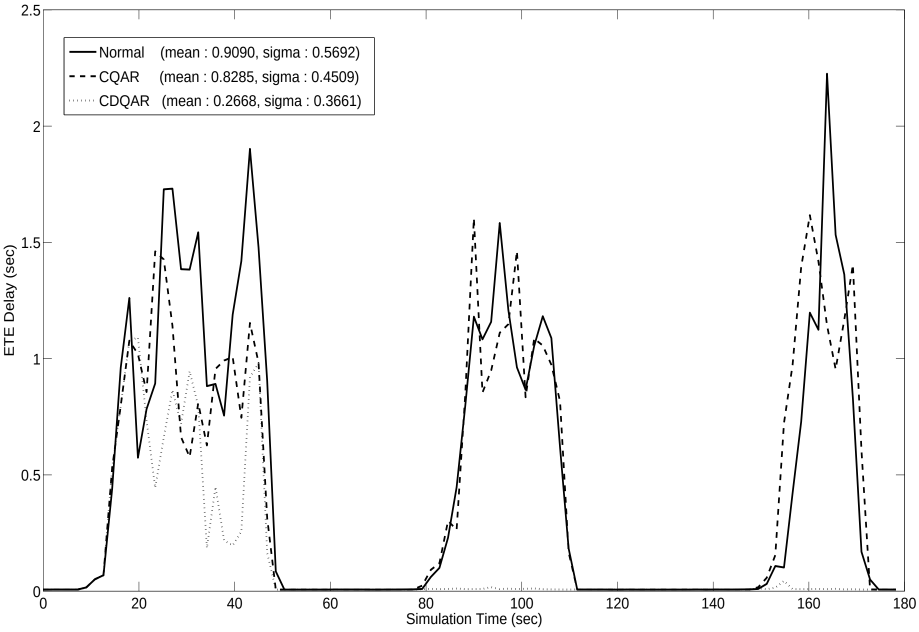


Figure 5.6. ETE delay of SS_{10} (downlink internet traffic)

30 % of the capacity of the queue, and both SSs use the alternative routes until there is a congestion in their distributed links. Thus, the same Internet traffic is sent using first two (in the first congestion), then one (in the second and third congestion) less centralized links and the system does not enter a congestion again. Since the second and third congestions are completely avoided the downlink ETE delays of all three-hop SSs revert to their uncongested values in these periods.

The effects of both methods can also be seen in the percentage of packet drops in Figures 5.7, 5.8, and 5.9 (Normalized by the total number of packets generated in the BS destined to SS_8, SS_9, and SS_{10} respectively). Similar to downlink ETE delays, the standard and both of our methods perform similarly regarding the uplink ETE delay values (Figures 5.10 and 5.11).

Both our methods work better than the normal routing method regarding the uplink ETE delay values (Figures 5.12, 5.13, and 5.14). CQAR decreases the delay values of SSs 8 and 10 considerably but because of the reason stated above, it gives instable results. Also, it cannot decrease the uplink ETE delay of SS_9, which has no

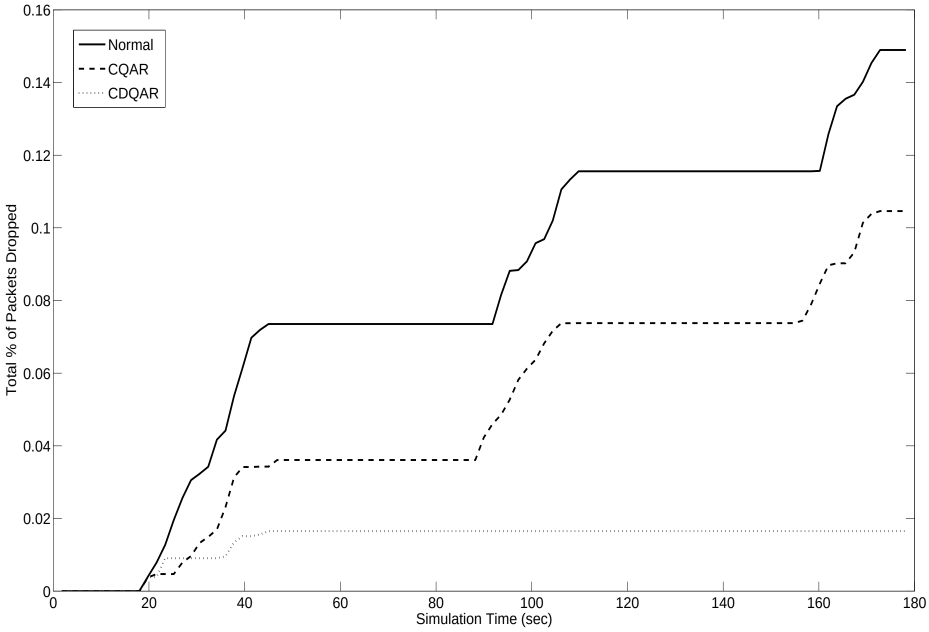


Figure 5.7. Total percentage of packets dropped SS_8 (downlink internet traffic)

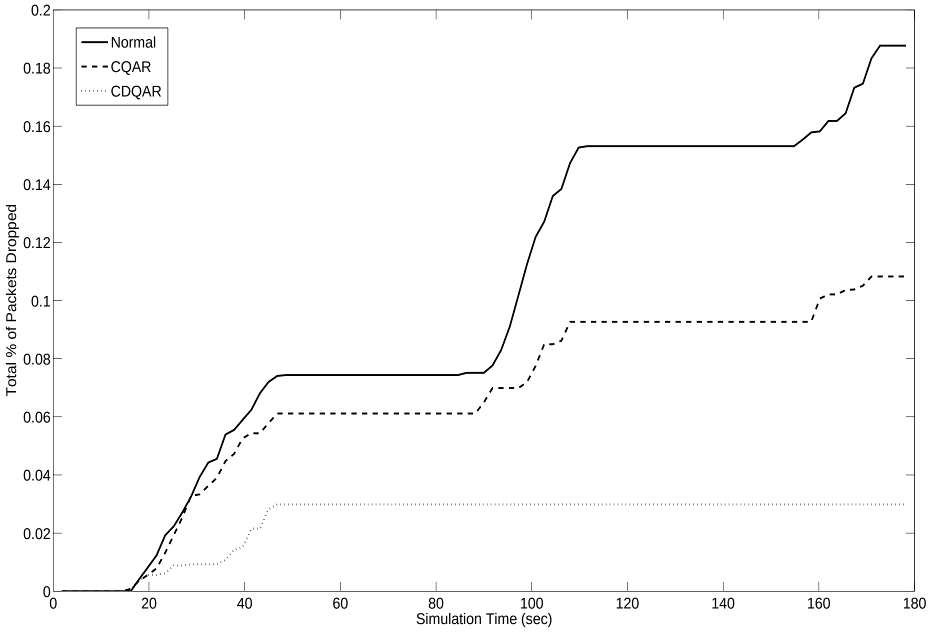


Figure 5.8. Total percentage of packets dropped SS_9 (downlink internet traffic)

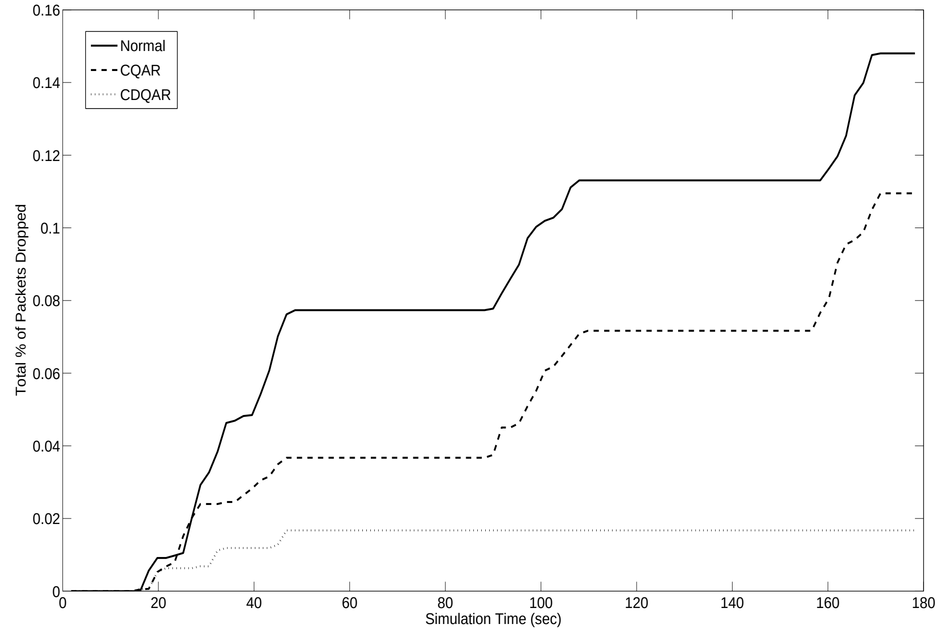


Figure 5.9. Total percentage of packets dropped SS_{10} (downlink internet traffic)

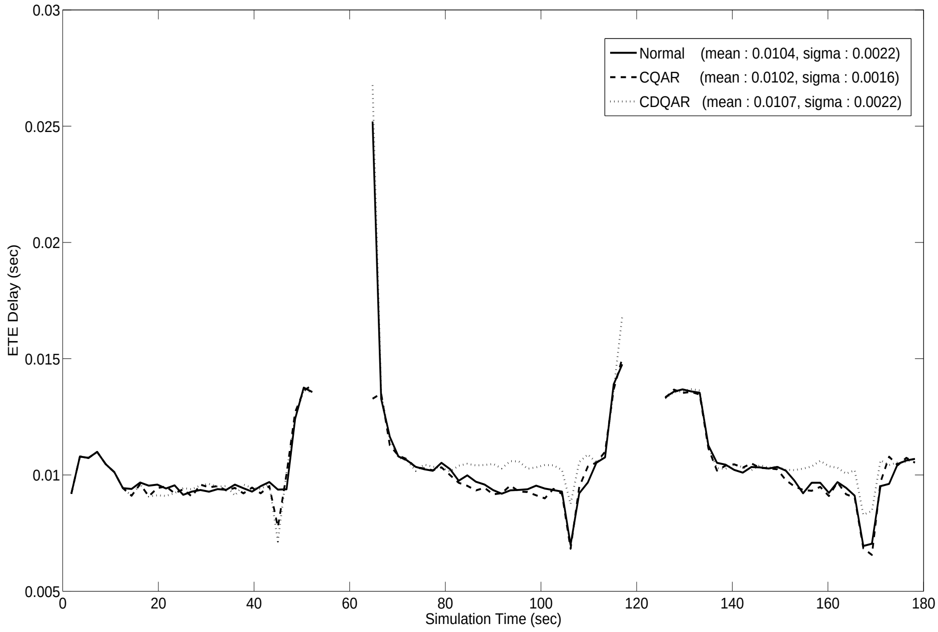


Figure 5.10. Average ETE delay of 1-hop SSs (uplink internet traffic)

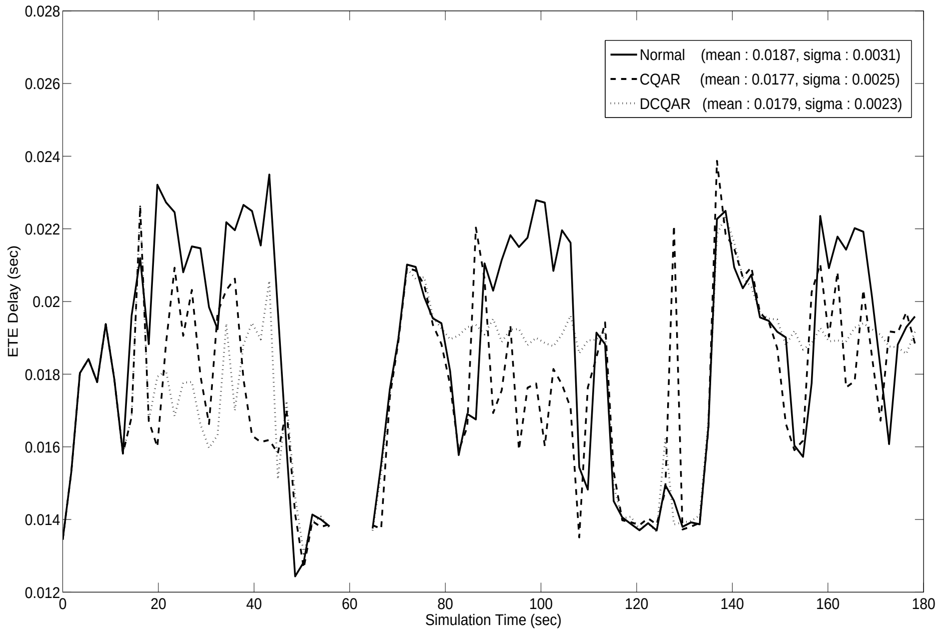


Figure 5.11. Average ETE delay of 2-hop SSs (uplink internet traffic)

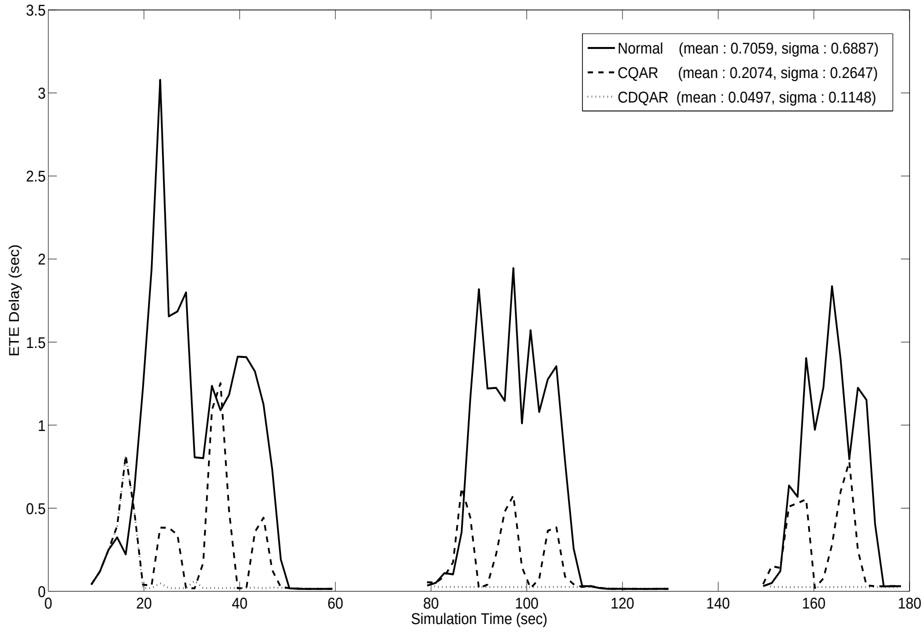


Figure 5.12. ETE delay of SS_8 (uplink internet traffic)

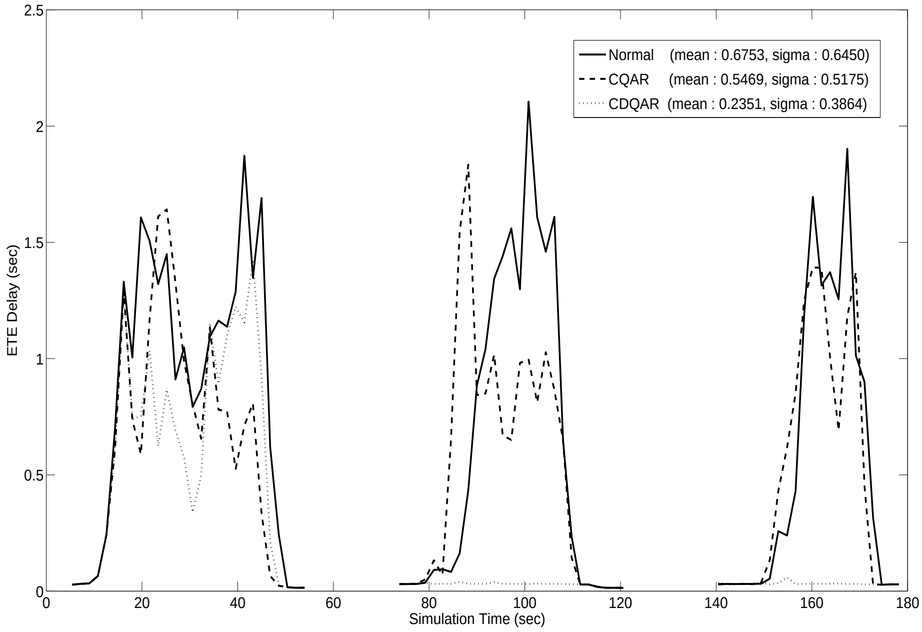


Figure 5.13. ETE delay of SS_9 (uplink internet traffic)

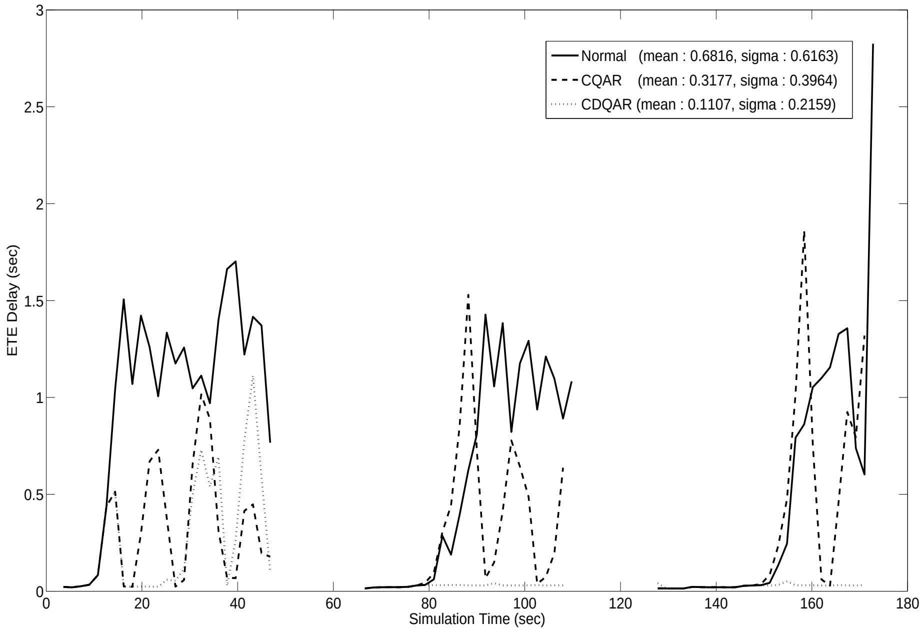


Figure 5.14. ETE delay of SS_{10} (uplink internet traffic)

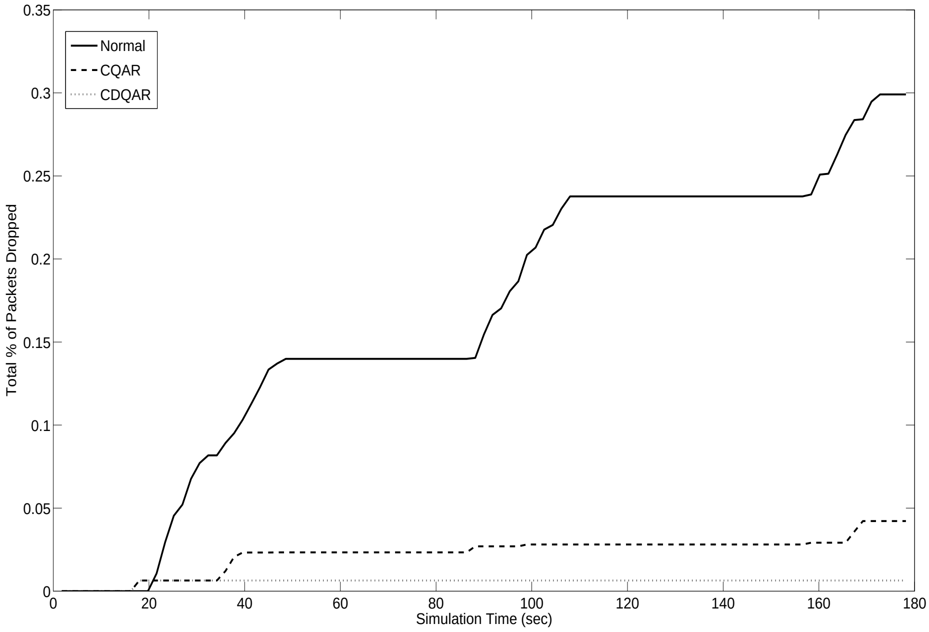


Figure 5.15. Total percentage of packets dropped SS_8 (uplink internet traffic)

pseudo-parent.

CDQAR, on the other hand, performs far more efficiently in terms of uplink ETE delay than CQAR. After the first route shifting, SS_8 always uses its pseudo-parent for its Internet traffic since it does not have any distributed traffic. However, SS_{10}, reverts back to using its parent node after 25th second when it starts generating intranet traffic to SS_9. Thus, its delay increases again after the 25th second and it performs worse than CQAR near the end of the first congestion period. But, since SS_8 does not revert back to its initial Internet route, the total load in the network is reduced and SS_{10}'s centralized link does not become congested in the second and third congestion periods and it does not need to use its pseudo-parent.

Unlike Internet traffic, our methods increase the ETE delay and number of dropped packets values (Figure 5.18). Since CQAR does not check the distributed queue lengths, it cannot know if there is intranet traffic. CDQAR again performs better than CQAR in this type of traffic. CDQAR checks distributed links and when it senses a high value in the distributed links is attained, it switches back to its original

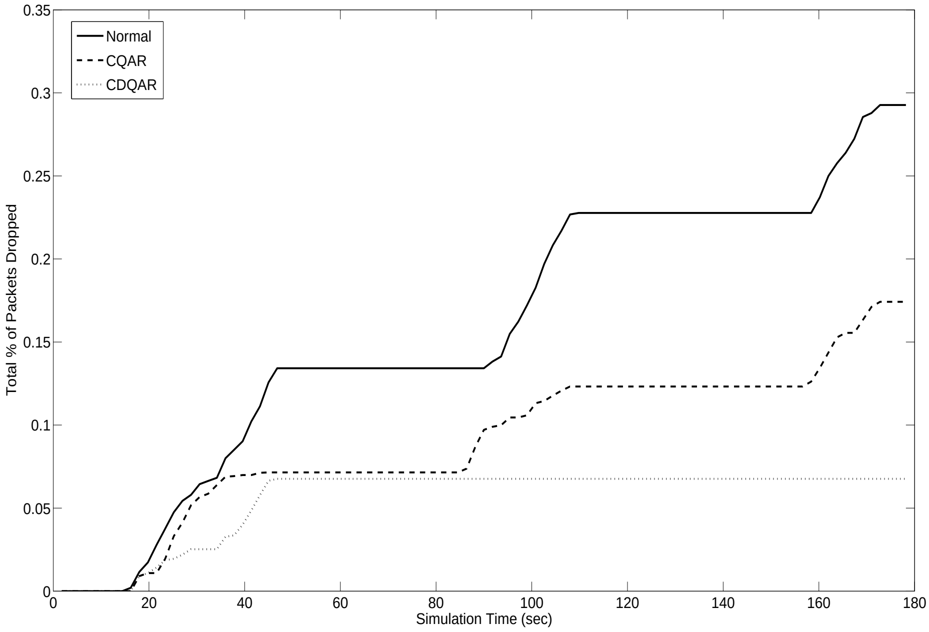


Figure 5.16. Total percentage of packets dropped SS_9 (uplink internet traffic)

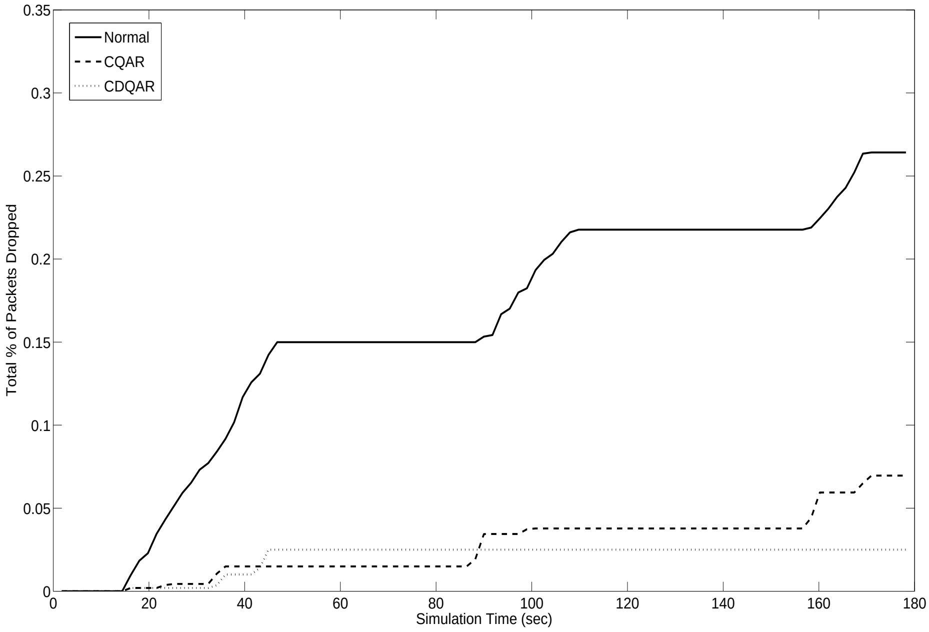


Figure 5.17. Total percentage of packets dropped SS_{10} (uplink internet traffic)

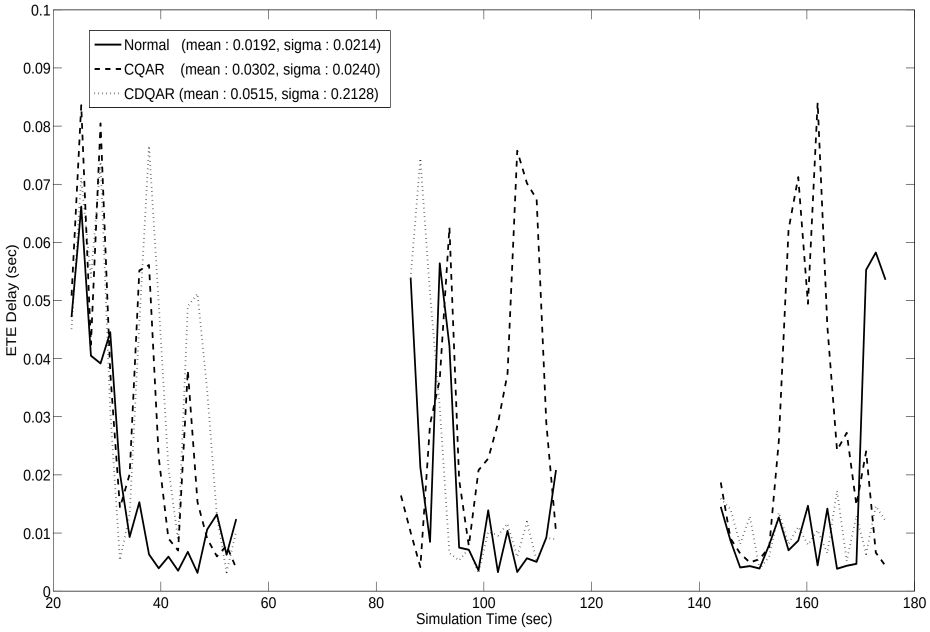


Figure 5.18. ETE delay (intranet Traffic SS_{10} - SS_{9})

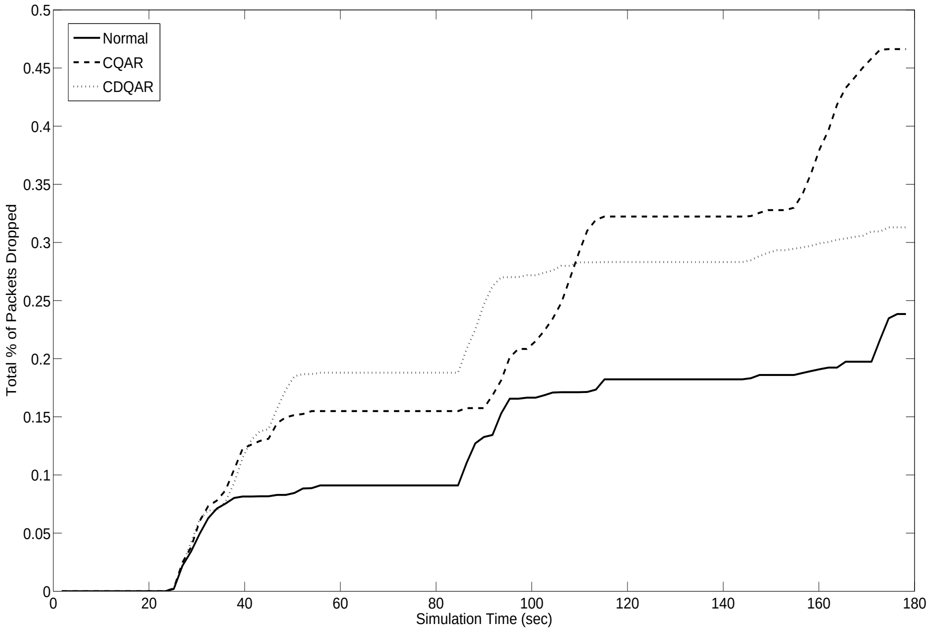


Figure 5.19. Total percentage of packets dropped (intranet Traffic SS_{10} - SS_{9})

route to hinder the intranet traffic. In the second and third congestion periods, since SS_8 does not need to use its pseudo-parent as stated above, the intranet traffic does not suffer any additional delay values.

6. CONCLUSION & FUTURE WORK

The Mesh operating mode of IEEE 802.16 adds intranet traffic capability and increased area coverage over the PMP operating mode of the standard. In the Mesh mode, the data subframe is divided into two parts for Internet and intranet traffics, respectively. We develop a cross layer routing scheme that uses links dedicated to intranet traffic for Internet traffic in case of a congestion in the Internet traffic. Since our method does not change any part of the original IEEE 802.16 standard, it can be implemented to both Mesh BS and SS devices easily.

Our scheme has been tested by two approaches. While the first one improves the systems carrying capacity and decrease the ETE delay of the congested node's and number of dropped packets, it gives instable results. The second approach gives much better results and provides a stable working state. This approach not only decreases the congested node's delay values but also decreases the ETE delay of other congested nodes since the routing decision affects the whole cell.

Other approaches can also be developed for better system capacity improvement, decision speed, and system stability. Since each node knows how much of the total Internet traffic part of the data subframe is available by the CSCH grant messages, a node can shift its Internet routing to its pseudo-parent when the system capacity is exceeded even if there is no congestion in its own traffic. This method, is a system-wise better approach than the two approaches we introduced in this thesis. As a future work we plan to implement this approach and compare its results with the results presented in this thesis.

REFERENCES

1. Kuran M. S., T. Tugcu, “A Survey on Emerging Broadband Wireless Access Technologies”, Elsevier Computer Networks, to be published in Vol. 51, Issue 11, pp. 3013 - 3046, Aug. 2007.

2. IEEE 802.16-2004, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems," Oct. 2004.

3. IEEE 802.16-2005, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems for Mobile Users," Dec. 2005.

4. Tanenbaum A. S., "Computer Networks 4th," Prentice Hall, 2003.

5. Cleary K., "Internet via MMDS," IEEE International Broadcasting Convention 1997 (IBC ’97), Amsterdam, Holland, 1997.

6. Nordbotten A., "LMDS Systems and their Application," IEEE Communications Magazine, Vol. 38, Issue 6, pp. 150 - 154, 2000.

7. Fong B., N. Ansari, A. C. M. Fong, G. Y. Hong, and P. B. Rapajic, "On the Scalability of Fixed Broadband Wireless Access Network Deployment," IEEE Radio Communications, Vol. 42, Issue 9, pp. s12 - s18, 2004.

8. Eklund C., R. B. Marks, K. L. Stanwood, and S. Wang, "IEEE Standard 802.16: A Technical Overview of the WirelessMAN Air Interface for Broadband Wireless Access," IEEE Communications Magazine, Vol. 40, Issue 6, pp. 98 - 107, Jun. 2002.

9. Cao M., W. Ma, Q. Zhang, X. Wang, and W. Zhu, "Modelling and Performance Analysis of the Distributed Scheduler in IEEE 802.16 Mesh Mode," ACM 6th International Symposium on Mobile Ad Hoc Networking and Computing 2005 (MobiHoc

'05), pp. 78 - 89, Boston, Massachusetts, USA, 2005.

10. Vaughan-Nichols S. J., "Achieving Wireless Broadband with WiMAX," IEEE Computer Magazine, Vol. 37, Issue 6, pp. 10 - 13, 2004.
11. Bellman R., "WiMAX: Broadband Wireless Access and Beyond," Business Communications Review, Vol. 8, Issue 11, pp. 46 - 53, 2004.
12. Ghosh A., D. R. Wolter, J. G. Andrews, and R. Chen, "Broadband Wireless Access with WiMAX/802.16: Current Performance Benchmarks and Future Potential," IEEE Communications Magazine, Vol. 43, Issue 2, pp. 129 - 136, 2005.
13. Hoymann C., "Analysis and Performance Evaluation of The OFDM-Based Metropolitan Area Network IEEE 802.16," Computer Network, Vol. 49, Issue 3, pp. 341 - 363, 2005.
14. WiMAX Forum, "Fixed, Nomadic, Portable and Mobile Applications for 802.16-2004 and 802.16e WiMAX Networks," http://www.WiMAXforum.org/news/downloads/Applications_for_802.16-2004_and_802.16e_WiMAX_networks_final.pdf, 2005.
15. Ramachandran S., C. W. Bostian, and S. F. Midkiff, "A Link Adaptation Algorithm for IEEE 802.16," IEEE Wireless Communications and Networking Conference 2005 (WCNC '05), Vol. 3, pp. 1466 - 1471, New Orleans, Los Angeles, USA, 2005.
16. Fu X. and H. Minn, "Initial Uplink Synchronization and Power Control (Ranging Process) for OFDMA Systems," IEEE Communications Society, 47th Annual IEEE Global Telecommunications Conference 2004 (Globecom '04), Vol. 6, pp. 3999 - 4003, 2004 Dallas, Texas, USA, 2004.
17. Shetiya H. and V. Sharma, "Algorithms for Routing and Centralized Scheduling in IEEE 802.16 Mesh Networks," IEEE Wireless Communications and Networking

Conference 2006 (WCNC ’06), Las Vegas, NV, USA, 2006.

18. Johnston D. and J. Walker, "Overview of IEEE 802.16 Security," IEEE Security & Privacy, Vol. 2, Issue 3, pp. 40 - 48, 2004.

19. Iyengar R., V. Sharma, K. Kar, and B. Sikdar, “Analysis of Contention-based Multi-channel Wireless MAC for Point-to-multipoint Networks,” IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2006 (WoWMoM ’06), pp. 453 - 455, Niagara-Falls, Buffalo-NY, USA, 2006.

20. Schwingenschlögl C., V. Dastis, P. S. Mogre, M. Hollick, and R. Steinmetz, “Performance Analysis of the Real-time Capabilities of Coordinated Centralized Scheduling in 802.16 Mesh Mode”, ??? 2006.

21. Leung K. K., S. Mukherjee, and G. E. Rittenhouse, “Mobility Support for IEEE 802.16d Wireless Networks,” IEEE Wireless Communications and Networking Conference 2005 (WCNC ’05), Vol. 3, pp. 1446 - 1452, New Orleans, Los Angeles, USA 2005.

22. Lawton G., “What Lies Ahead for Cellular Technology?,” IEEE Computer, Vol. 38, Issue 6, pp. 14 - 17, 2005.

23. Lee D. H., K. Kyamakya, and J. P. Umodi, “Fast Handover Algorithm for IEEE 802.16e Broadband Wireless Access System,” 1st International Symposium on Wireless Pervasive Computing 2006 (ISWPC ’06), pp. 1 -6, Phuket, Thailand, 2006.

24. Seo J. B., S. Q. Lee, N. H. Park, H. W. Lee, and C. H. Cho, “Performance Analysis of Sleep Mode Operation in IEEE802.16e,” IEEE 60th Vehicular Technology Conference 2004 (VTC ’04), Vol. 2, pp. 1169 - 1173, Los Angeles, CA, USA 2004.

25. Xiao Y., “Energy Saving Mechanism in the IEEE 802.16e Wireless MAN,” IEEE Communications Letters, Vol. 9, Issue 7, pp. 595 - 597, 2005.

26. Zhang Y. and M. Fujise, “Energy Management in the IEEE 802.16e MAC,” IEEE Communications Letters, Vol. 10, Issue 4, pp. 311 - 313, 2006.D. Johnson, C. Perkins and J. Arkko, "Mobility Support in IPv6", RFC 3775, IETF, June 2004.

27. Lee M. J., J. Zheng, Y-B. Ko, and D. M. Shrestha, “Emerging Standards for Wireless Mesh Technology,” IEEE Wireless Communications, Vol. 13, Issue 2, pp. 56 - 63, 2006.

28. Lee H., T. Kwon, and D. Cho, “An Efficient Uplink Scheduling Algorithm for VoIP Services in IEEE 802.16 BWA Systems,” IEEE 60th Vehicular Technology Conference 2004 (VTC ’04), Vol. 5, pp. 3070 - 3074, Los Angeles, CA, USA, 2004.

29. Cicconetti C., L. Lenzini, E. Mingozzi, and C. Eklund, "Quality of Service Support in IEEE 802.16 Networks," IEEE Network, Vol. 20, Issue 2, pp. 50 - 55, 2006.

30. Chu G., D. Wang and S. 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 2002 (ICCCAS ’02), Vol. 1, pp.435 - 439, Chengdu, China, 2002.

31. Wongthavarawat K. and A. Ganz, “IEEE 802.16 Based Last Mile Broadband Wireless Military Networks with Quality of Service Support,” IEEE Military Communications Conference 2003 (MILCOM ’03), Vol. 2, pp. 779 - 784, Monterey, CA, USA, 2003.

32. Jiang C-H. and T-C. Tsai, “Token Bucket Based CAC and Packet Scheduler for IEEE 802.16 Broadband Wireless Access Networks,” IEEE 3rd Consumer Communications and Networking Conference 2006 (CCNC ’06), Vol. 1, pp. 183 - 187, Las Vegas, Nevada, USA, 2006.

33. Niyato D. and E. Hossain, “Queue-Aware Uplink Bandwidth Allocation and Rate Control for Polling Service in IEEE 802.16 Broadband Wireless Networks,” IEEE Transactions on Mobile Computing, Vol. 5, Issue 6, pp. 668 - 679, 2006.

34. Kim D. and A. Ganz, “Fair and Efficient Multihop Scheduling Algorithm for IEEE 802.16 BWA Systems,” International Conference on Broadband Networks 2005 (BROADNETS ’05), Vol. 2, pp. 833- 839, Boston, MA, USA, 2005.

35. Cheng S-M., P. Lin, D-W. Huang, and S-R. Yang, “A Study on Distributed/Centralized Scheduling for Wireless Mesh Network,” International Wireless Communications and Mobile Computing Conference 2006 (IWCMC ’06), pp. 599 - 604, Vancouver, British Columbia, Canada, 2006.