2

A Comprehensive Survey on WiMAX Scheduling Approaches

Lamia Chaari, Ahlem Saddoud,
Rihab Maaloul and Lotfi Kamoun
*Electronics and Information Technology Laboratory,
National School of Engineering of Sfax (ENIS),
Tunisia*

1. Introduction

The institute of Electrical and Electronics IEEE 802.16 standard is a real revolution in wireless metropolitan area networks (wireless MANs) that enables high-speed access to data, video and voice services. The IEEE 802.16 is mainly aimed at providing broadband wireless access (BWA). Thus, it complements existing last mile wired networks such as cable modem and xDSL. Its main advantage is fast deployment which results in cost saving.

WiMAX networks are providing a crucial element in order to satisfy on-demand media with high data rates. This element is the QoS and service classes per application. In Broadband Wireless communications, QoS is still an important criterion. So the basic feature of WiMAX network is the guarantee of QoS for different service flows with diverse QoS requirements. While extensive bandwidth allocation and QoS mechanisms are provided, the details of scheduling and reservation management are left not standardized. In fact, the standard supports scheduling only for fixed-size real-time service flows. The scheduling of both variable-size real-time and non-real-time connections is not considered in the standard. Thus, WiMAX QoS is still an open field of research and development for both constructors and academic researchers. The standard should also maintain connections for users and guarantee a certain level of QoS. Scheduling is the key model in computer multiprocessing operating system. It is the way in which processes are designed priorities in a queue. Scheduling algorithms provide mechanism for bandwidth allocation and multiplexing at the packet level.

In this chapter, we proposed a survey on WiMAX scheduling scheme in both uplink and downlink traffic. The remainder of this chapter is organized as follows: Section 2 presents the QoS support in WiMAX networks, and section 3 presents scheduling mechanisms classifications. In section 4, we discuss channel-unaware and channel aware schedulers proposed for both uplink and downlink. We present the relay WiMAX schedulers in section 5. Section 6 presents a comparative study. Finally, we conclude the chapter in section 7.

2. Quality of services provisioning in WiMAX networks

2.1 Services and parameters

In WiMAX (Jeffrey,2007)(Labiod & Afifi, 2007)(Shepard,2006)(Nuaymi, 2007), a service flow is a MAC transport service provided for transmission of uplink, downlink traffic, and is a key concept of the QoS architecture. Each service flow is associated with a unique set of QoS parameters, such as latency, jitter throughput, and packet error rate. The various service flows admitted in a WiMAX network are usually grouped into service flow classes, each identified by a unique set of QoS requirements. This concept of service flow classes allows higher-layer entities at the subscriber station (SS) and the base station (BS) to request QoS parameters in globally consistent ways. The WiMAX networks is a connection-oriented MAC in that it assigns traffic to a service flow and maps it to MAC connection using a Connection ID (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. A service flow is a unidirectional flow of packets that is provided a particular QoS. The SS and BS provide this QoS according to the QoS parameter set defined for the service flow. Each data service is associated with a set of QoS parameters that quantify its behavior aspects. These parameters are managed through a series of MAC management messages referred to as DSA, DSC, and DSD. The DSA messages create a new service flow. The DSC messages change an existing service flow. The DSD messages delete an existing service flow. An SS wishing to either create an uplink or downlink service flow sends a request to the BS using a DSA-REQ message. The BS checks the integrity of the message and, if the message is intact, sends a message received (DSX-RVD) response to the SS. The BS checks the SS's authorization for the requested service and whether the QoS requirements can be supported, generating an appropriate response using a DSA-RSP message. The SS concludes the transaction with an acknowledgment message (DSA-ACK). An SS that needs to change a service flow definition performs the following operations. The SS informs the BS using a DSC-REQ. The BS checks the integrity of the message and, if the message is intact, sends a message received (DSX-RVD) response to the SS. The BS shall decide if the referenced service flow can support this modification. The BS shall respond with a DSC-RSP indicating acceptance or rejection. In the case when rejection was caused by presence of non-supported parameter of non-supported value, specific parameter may be included into DSC-RSP. The SS reconfigures the service flow if appropriate, and then shall respond with a DSC-ACK. Any service flow can be deleted with the DSD messages. When a service flow is deleted, all resources associated with it are released. 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.

Five services are supported in the mobile version of WiMAX: Unsolicited Grant Service (UGS), Real-Time Polling Service (rtPS), Extended Real-Time Polling Service (ErtPS) , non-real-time polling service (nrtPS), and Best Effort (BE). Each of these scheduling services has a mandatory set of QoS parameters that must be included in the service flow definition when the scheduling service is enabled for a service flow. These are summarized in Table 1.

QoS Category	Applications	QoS Specifications
UGS Unsolicited Grant Service	VoIP	-Maximum Sustained Rate -Maximum Latency Tolerance -Jitter Tolerance
rtPS Real-Time Polling Service	Streaming Audio or Video	-Minimum Reserved Rate -Maximum Sustained Rate -Maximum Latency Tolerance -Traffic Priority
ErtPS Extended Real-Time Polling Service	Voice with Activity Detection (VoIP)	-Minimum Reserved Rate -Maximum Sustained Rate -Maximum Latency Tolerance -Jitter Tolerance -Traffic Priority
nrtPS Non-Real-Time Polling Service	File Transfer Protocol (FTP)	-Minimum Reserved Rate -Maximum Sustained Rate -Traffic Priority
BE Best-Effort Service	Data Transfer, Browsing, Web etc.	-Maximum Sustained Rate -Traffic Priority

Table 1. WiMAX applications and QoS specifications

2.2 Functional elements

Based on the IEEE 802.16e specification (Standard, 2006), the proposed QoS functional elements includes call admission control (CAC), scheduling and bandwidth allocation.

2.2.1 Bandwidth allocation schemes

During initialization and network entry, the BS assigns up to three dedicated CID to each SS in order to provide the SS the ability to sends and receives control messages. The SS can send the bandwidth request message to the BS by numerous methods. In the IEEE 802.16 standard, bandwidth requests are normally transmitted in two modes: a contention mode and a contention-free mode (polling). In the contention mode, the SSs send bandwidth-requests during a contention period, and the BS using an exponential back-off strategy resolves contention. In the contention-free mode, the BS polls each SS, and an SS in reply sends its BW-request. The basic intention of unicast polling is to give the SS a contention-free opportunity to tell the BS that it needs bandwidth for one or more connections In addition to polling individual SSs, the BS may issue a broadcast poll by allocating a request interval to the broadcast CID, when there is insufficient bandwidth to poll the stations individually.

Similarly, the standard provides a protocol for forming multicast groups to give finer control to contention-based polling. SSs with currently active UGS connections may set the

PM bit (bit PM in the Grant Management subheader) in a MAC packet of the UGS connection to indicate to the BS that they need to be polled to request bandwidth for non-UGS connections. Variable bandwidth assignment is possible in rtPS, nrtPS and BE services, whereas UGS service needs fixed and dedicated bandwidth assignment. The BS periodically in a fixed pattern offers bandwidth for UGS connections so UGS connections do not request bandwidth from the BS. 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. 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 SSs needs.

Furthermore, the IEEE 802.16 MAC accommodates two classes of SS, differentiated by their ability to accept bandwidth grants simply for a connection or for the SS as a whole. Both classes of SS request bandwidth per connection to allow the BS uplink-scheduling algorithm to properly consider QoS when allocating bandwidth. With the grant per connection (GPC) class of SS, bandwidth is granted explicitly to a connection, and the SS uses the grant only for that connection. With the grant per SS (GPSS) class, SSs are granted bandwidth aggregated into a single grant to the SS itself. GPC is more suitable for few users per subscriber station. It has higher overhead, but allows a simpler SS 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.

Based on the methods by which the SS can send the bandwidth request message to the BS, bandwidth allocation mechanisms can be classified according table 2.

2.2.2 Call Admission Control

Researchers have characterized CAC as the decision maker for the network. When a subscriber station SS send a request to the base station (BS) with a certain QoS parameters for a new connection, the BS will check whether it can provide the required QoS for that connection. If the request was accepted, the BS verifies whether the QoS of all the ongoing connections can be maintained. Based on this it will take a decision on whether to accept or reject the connection. The process described above is called as CAC mechanism. The basic components in an admission controller are performance estimator which is used to obtain the current state of the system; resource allocator uses this state to reallocate available radio resource. Then the admission control decision is made to accept or reject an incoming connection. A connection is admitted: if there is enough bandwidth to accommodate the new connection. The newly admitted connection will receive QoS guarantees in terms of both bandwidth and delay and QoS of existing connections must be maintained (Chou et al,2006). A more relaxed rule would be considered to limit admission control decision (to reject) to applications with real-time hard constraints, for example, IP telephony and video conferencing. For other requests (e.g: video streaming, web browsing) if there are insufficient resources, one can provide throughput less than requested by them. A simple admission control decision can be evident: if there are enough available resources in the BS so new connections are admitted else it will be rejected. However, a simple admission

Type	QoS Classes	Mechanisms
Unsolicited request	UGS and ertPS	- Periodically allocates bandwidth at setup stage: - No overhead and meet guaranteed latency for real-time service - Exhausted bandwidth if it is granted and the flow has no packets to send.
Poll-me bit (PM)	UGS	-Asks BS to poll non UGS connections implicitly in MAC header No overhead but Still needs the unicast polling
Piggybacking	ertPS, rtPS, BE & nrtPS	- Piggyback BWR over any other MAC packets being sent to the BS. - Do not need to wait for poll, Less overhead; 2 bytes vs. 6 bytes Grant management.
Bandwidth stealing	nrtPS and BE	- Sends BWR instead of general MAC packet - BWR (6 bytes = MAC header) - Do not need to wait for poll
Contention region (WiMAX)	ertPS, nrtPS and BE	- MSs use contention regions to send BWR ➔ Need the backoff mechanism - Overhead Adjustable - Reduced polling overhead
Codeword over CQICH	ertPS	- Specifies codeword over CQICH - Makes use of CQI channel - Limit number of bandwidth on CQICH
CDMA code-based BWR (Mobile WiMAX)	nrtPS and BE	- MS chooses one of the CDMA request codes from those set aside for bandwidth requests. - Six sub channels over 1 OFDM symbol for up to 256 codes - Reduced polling overhead compared to contention region - Results in one more frame delay compared to contention region
Unicast Polling	ertPS, rtPS, nrtPS and BE	- BS polls each MS individually and periodically. - Guarantees that MS has a chance to ask for bandwidth - More overhead BWR (6 bytes per MS) periodically
Multicast, Broadcast and Group Polling	ertPS, nrtPS and BE	- BS polls a multicast group of MSs. - BWR (6 bytes) per multicast ➔ Reduced polling overhead - Some MSs may not get a chance to request bandwidth; need contention resolution technique.

Table 2. Taxonomy of Bandwidth request mechanisms

control is not efficient to guarantee QoS of different types of connections and in the same time, it can affect the performance of IEEE 802.16 network. An important question might be asked: What is the decision of the call admission module when no resources are available for new flows? The answer must be a solution to avoid dropping and blocking new connection requests when it is possible. These solutions are presented in the proposals described below. We present a classification and a description of CAC algorithms proposed in the literature for PMP (Point-to- Multipoint) mode. We classify CAC proposals into two classes. The first category is called "*with degradation*"; it is based on decreasing the resources provided to existing connections in the purpose to allow a new service flow to be accepted in the network. In the second policy named "*without degradation*", it is forbidden to adopt any strategy of degradation in order to maintain the QoS of existing connections. Figure1 shows a diagram with the topics used in the classification.

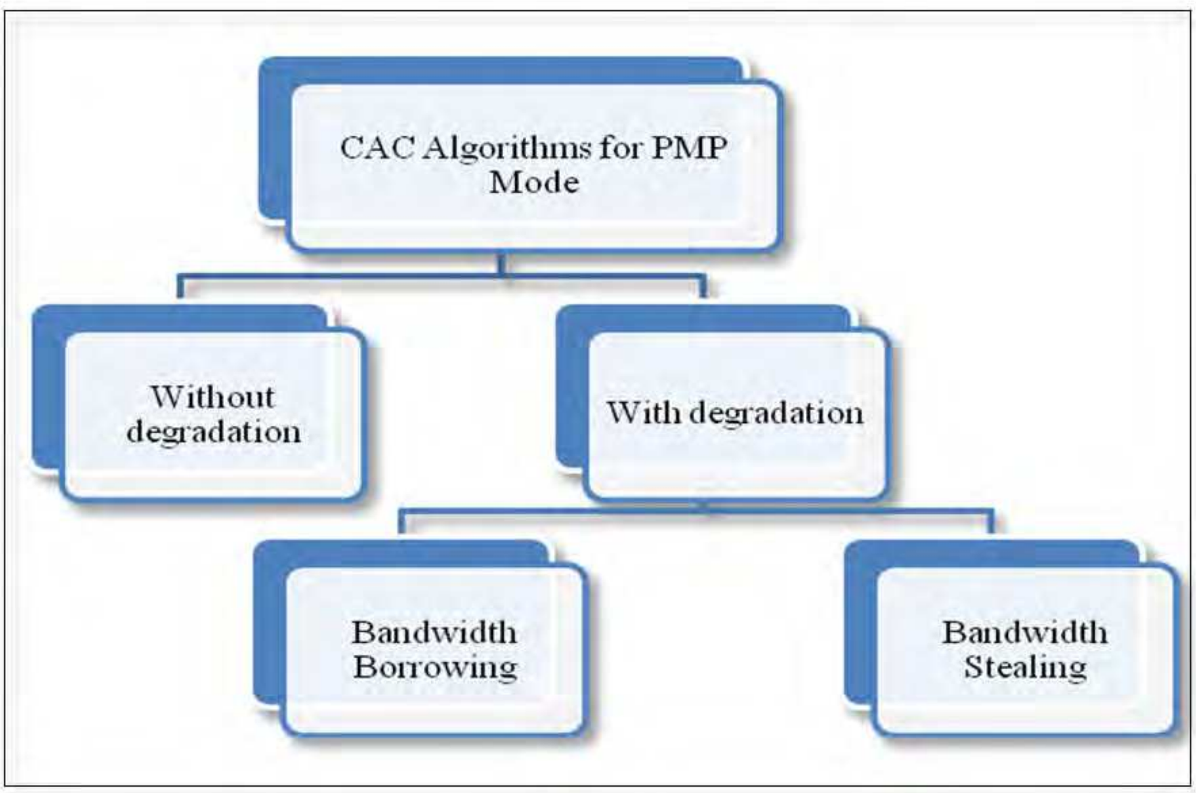


Fig. 1. Proposed classification for WiMAX CAC algorithms in PMP mode

First, "*without degradation*" policy is more flexible than the second one as it offers more opportunities and chance for new requests to be accepted and to get the possible resources when it is necessary. Second, CAC schemes based on degradation strategies are unfortunately less conservative and not simple.

We classify the CAC scheme based on degradation policy in two sub-classes: based on bandwidth borrowing mechanism, or bandwidth stealing. The main concept of these CAC schemes is to decrease the resources afforded to existing connections in order to support requests of a new service flow and to satisfy their demand.

We have regrouped and compared the most related CAC proposals in table 3.

Proposals	QoS Parameter (min/max)	Analytical validation	Token Bucket policy	Bandwidth estimation	Degradation strategy
(H.Wang et al, 2005)	Max Bandwidth Utilization	Markov	N	N	borrowing
(Zhu & Lu, 2006)	Max Bandwidth Utilization	Markov	N	N	borrowing
(Kalikivayi et al, 2008)	Delay guarantee	Markov	S	S	borrowing
(Kitti & Aura, 2003)	Delay guarantee	N	S	N	N
(Wang et al, 2007)	Min blocking probability	N	N	N	borrowing
(Tzu-Chieh et al, 2006)	Delay guarantee	Markov chain	S	N	stealing
(Shafaq et al, 2007)	Min blocking probability	N	N	S	N
(Chandra & Sahoo, 2007)	Delay & jitter	N	N	S	N
(Yu et al, 2009)	Delay, Min blocking probability	Markov chain	N	N	N
(Rango et al, 2011)	throughput, average delay and jitter	N	N	S	N
(Shida & Zisu, 2008)	Max throughput of all flows and decrease the delay of the VBR	N	N	S	N

N: Not supported S: supported

Table 3. CAC in IEEE 802.16 PMP Mode: A Comparative table

An admission control module in BS (J.Chen et al, 2005a) (Carlos,2009) has as input a Dynamic Service Addition (DSA; essentially a new connection), Dynamic Service Change (DSC) or a Dynamic Service Deletion (DSD) requests, either. These need to be considered in terms of a set of predefined QoS parameters. It also needs to know the current resource state of the network, which it can only determine by consulting the Scheduler. With that information, it applies the particular CAC algorithm and informs the scheduler of whether a request has been admitted or not. Most of the scheduling algorithm presented in literature assumes a simple CAC is present but this is inappropriate in some cases. Since both CAC and scheduling handle, the QoS a proper CAC algorithm is needed in order to guarantee the promised QoS. Sometimes CAC and scheduling algorithm working on different criteria can interfere, which necessitate CAC algorithms that works in an independent manner from the scheduling algorithm based on bandwidth and delay prediction (Castrucci et al,2008).

2.2.3 MAC scheduling services

In WiMAX network, a service flow is a MAC transport service provided for transmission of uplink, downlink traffic, and is a key concept of the QoS architecture. Each service flow is associated with a unique set of QoS parameters, such as latency, jitter throughput, and packet error rate. The various service flows admitted in a Mobile WiMAX network are usually grouped into service flow classes. This concept of service flow classes allows higher-layer entities at the SS and the BS to request QoS parameters in globally consistent ways. A service flow is a unidirectional flow of packets that is provided a particular QoS. The SS and BS provide this QoS according to the QoS Parameter Set defined for the service flow.

A service flow is partially characterized by the following attributes: (Standard, 2004)

- Service Flow ID: An SFID is assigned to each existing service flow. The SFID serves as the principal identifier for the service flow in the network. A service flow has at least an SFID and an associated direction. The SFID identifies a services which in turn identifies the right of the IEEE 802.16 SS to certain system resources, and also defines which of user's packets will be mapped to the corresponding MAC connection
- CID: Mapping to an SFID that exists only when the connection has an admitted or active service flow.
- "ProvisionedQoSParamSet": A QoS parameter set provisioned: When a service level was set up (neither reserved nor allocated).
- "AdmittedQoSParamSet": Defines a set of QoS parameters for which the BS (and possibly the SS) is reserving resources. The principal resource to be reserved is bandwidth.
- "ActiveQoSParamSet": Defines a set of QoS parameters defining the service actually being provided to the service flow. Only an Active service flow may forward packets.
- Authorization Module: A logical function within the BS that approves or denies every change to QoS Parameters and Classifiers associated with a service flow.

Scheduling is the main component of the MAC layer that assures QoS to various service classes. The MAC scheduling Services are adopted to determine which packet will be served first in a specific queue to guarantee its QoS requirement. In fact, the scheduler works as a distributor in order to allocate the available resources among SSs. Thus, an efficient scheduling algorithm could enhance the QoS provided by IEEE 802.16 network. As well, scheduling architecture should ensure good use of bandwidth, maintain the fairness among users, and satisfy the requirements of QoS. It is important to mention that Scheduling algorithms can be implemented in the BS as well as in the SSs. Those are implemented at the

BS have to deal with both uplink and downlink traffics. Therefore, there are three different schedulers: two at the BS schedule the packet transmission in downlink and uplink sub frame and the latter at the SS for uplink to apportion the assigned BW to its connections.

In order to indicate the allocation of transmission intervals in both uplink and downlink, in each frame, the signaling messages UL-MAP and DL-MAP are broadcasted at the beginning of the downlink sub frame. The scheduling decision for the downlink traffic is relatively simple as only the BS transmits during the downlink sub frame and the queue information is located in the BS. While, an uplink scheduler at the BS must synchronize its decision with all the SSs.

We describe a better understanding of some specific factors that should be considered in the scheduling policy as follows:

- **QoS requirements**: An efficient scheduling algorithm could enhance the QoS specification of the different types of service classes as it is mentioned in table1.
- **Fairness**: Besides assuring the QoS requirements, the bandwidth resources should be shared fairly between users. Thus, fairness represents one of the most challenging problems in the scheduling approaches.
- **Channel Utilization**: It is the fraction of time used to transmit data packets. It is almost equal to the channel capacity in PMP communications. A scheduling mechanism has to check that resources are not allocated to SSs that do not have enough data to send, thus resulting in wastage of resources.
- **Complexity**: The scheduling algorithm must be simple, fast and should not have a prohibitive implementation complexity as it serves different service classes in various constraints.
- **Scalability**: It is the capacity to handle a growing number of flows. A scheduling algorithm should efficiently operate as the number of connections increases.
- **Cross-layer design**: A scheduling algorithm should take into account the characteristics of different layers (e.g. the adaptive modulation and coding (AMC) scheme). It is significant to consider the burst profile in such scheduling policy in order to improve system performance.

2.3 A QoS framework

A novel design paradigm, the so-called cross-layer optimization, is one of the most promising issues of research for the improvement of wireless communication systems (Zhang & Chen, 2008). Cross-layer operation can be formulated conceptually as the selection of strategies across multiple layers such that the resultant interlayer operation is optimized. Each layer has optimal schemes under given states, such as channel condition and QoS parameters, and the combination of schemes selected in all layers results in optimized interlayer operation. In this section, we elaborate architecture for integrated QoS control with respect to cross-layer design. The IEEE 802.16 uses the PMP centralized MAC architecture where the BS scheduler controls all the system parameters (radio interface). It is the role of the BS scheduler to determine the burst profile and the transmission periods for each connection; the choice of the coding and modulation parameters are decisions that are taken by the BS scheduler according to the quality of the link and the network load and demand. Therefore, the BS scheduler must monitor the received carrier-to-interference-plus-noise-ratio (CINR) values (of the different links) and then determine the bandwidth requirements of each station taking into consideration the service class for this connection and the quantity of traffic required. Figure2 shows the BS scheduler operation based on cross layer approach.

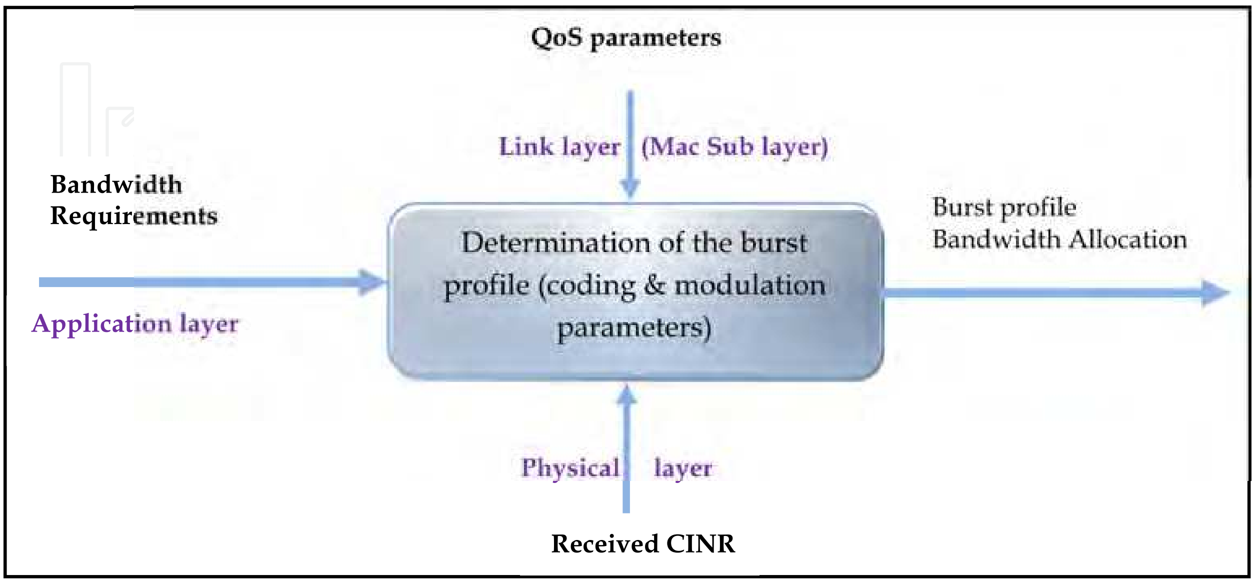


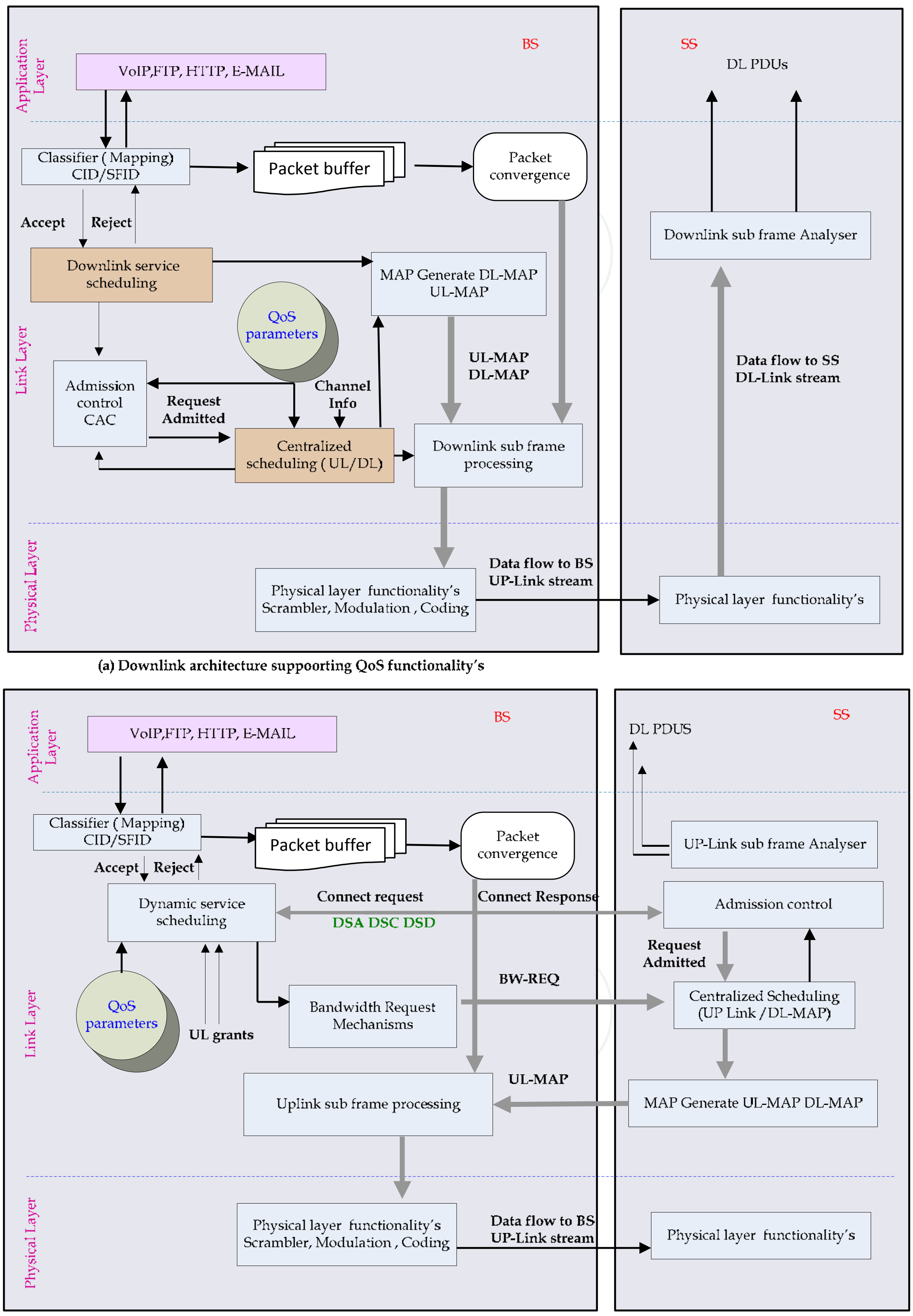
Fig. 2. Burst profile parameter

In figure 3, we give an idea about the architecture of the IEEE 802.16 QoS platform of the BS and SSs to support multimedia services.

This chapter emphasis especially in relationship between modules and the control information flows to provide cross-layer operation. In the downlink, all decisions related to the allocation of bandwidth to various SSs are made by the BS on a per CID basis. As MAC PDUs arrive for each CID, the BS schedules them for the PHY resources, based on their QoS requirements. Once dedicated PHY resources have been allocated for the transmission of the MAC PDU, the BS indicates this allocation to the SS, using the DL-MAP message. While the scheduler independently builds the DL-MAP and UL-MAP, the CAC needs to closely consult these in order to determine the available resources and consequently, whether to admit or deny a connection of a particular traffic type. Frames arriving at the BS were previously scheduled on the UL-MAP to be either BW requests or data PDUs to be forwarded on the DL or data PDUs destined for the BS itself. A BW request must be taken up by the CAC that decides whether to admit the request and, if so, will pass this information to the centralized scheduler.

The UL packets from the upper layer are classified into service flows by a packet classifier within the SS, and the SS requests BW according to the UL grant/scheduling type. From the amount of BW requested, the BS estimates the queue status information of each SS. In IEEE 802.16 systems, all resources are managed by the BS, thus the BS performs channel- and QoS-aware scheduling, on the basis of measured UL channel information, the negotiated QoS parameter and estimated queue status.

In the uplink, the SS requests resources by either using a stand-alone bandwidth-request MAC PDU or piggybacking bandwidth requests on a generic MAC PDU, in which case it



(a) Downlink architecture suppoorting QoS functionality's

(b) Uplink architecture suppoorting QoS functionality's

Fig. 3. QoS support for multimedia services in IEEE 802.16

uses a grant-management sub header. Since the burst profile associated with a CID can change dynamically, all resource requests are made in terms of bytes of information, rather than PHY layer resources, such as number of sub channels and/or number of OFDM symbols.

Each SS to BS (uplink) connection is assigned a scheduling service type as part of its creation. When packets are classified in the Convergence Sublayer (CS), the connection into which they are placed is chosen based on the type of QoS guarantees that are required by the application.

Service flows may be created, changed or deleted. This is accomplished through a series of MAC management messages: DSA, DSC and DSD. The DCD/UCD (Downlink/Uplink Channel Descriptor) message are broadcasted MAC management message transmitted by the BS at a periodic time interval in order to provide the burst profiles (physical parameter sets) that can be used by a downlink/Uplink physical channel during a burst.

As shown in figure 3 the most important QoS modules are the uplink scheduler (SS), the centralized scheduler (BS) and the downlink scheduler (BS), so the scheduling architectures of those modules implementation are illustrated in figure 4.

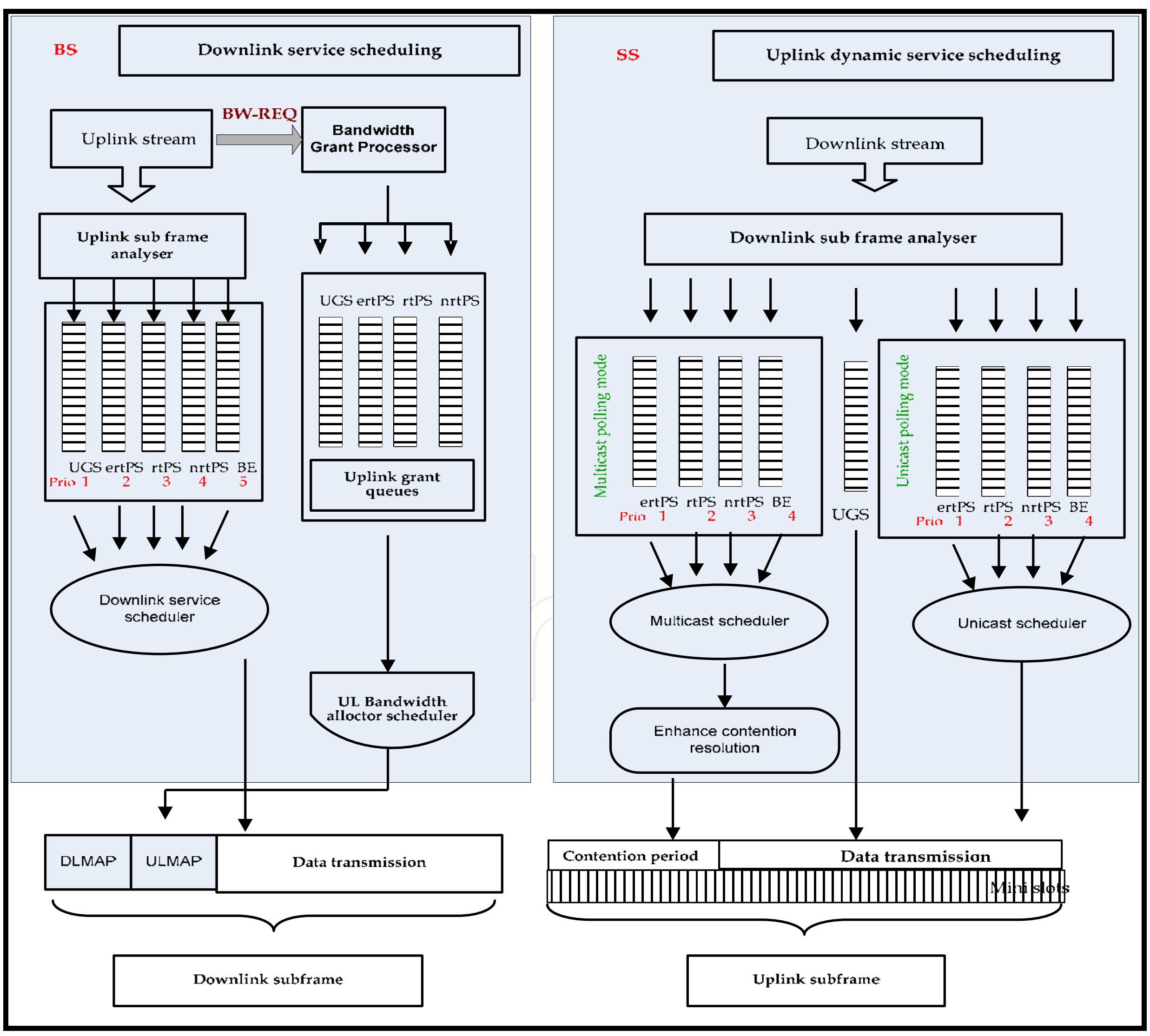


Fig. 4. Scheduling architecture in BS and SS using TDD mode

The WiMAX MAC layer uses a scheduling service to deliver and handle SDUs and MAC PDUs with different QoS requirements. A scheduling service uniquely determines the mechanism the network uses to allocate UL and DL transmission opportunities for the PDUs. When packets are classified in the convergence sublayer, the connection into which they are placed is chosen based on the type of QoS guarantees that are required by the application.

3. Scheduling mechanisms classification

In the research literature, we find an important number of studies focus on mechanisms for packet scheduling in WiMAX networks (Kitti & Aura, 2003)(Sonia & Hamid, 2010)(Ridong et al, 2009)(G.Wei et al, 2009). We classify the scheduling methods proposed in the literature of IEEE 802.16 as is shown in figure 5. The scheduling algorithms used in WiMAX network could be originally designed for wired network in order to satisfy the QoS requirements. Therefore, these algorithms do not take into account WiMAX channel characteristics. The Schedulers of this kind is belonging to the channel unaware scheduling category. But the scheduling algorithm which takes into account the variability of channel characteristics can be categorized as channel aware scheduler. The objective of the following sections is to provide a comprehensive survey on the scheduling research works proposed for WiMAX. These works are described according to the above taxonomy illustrated by the figure5.

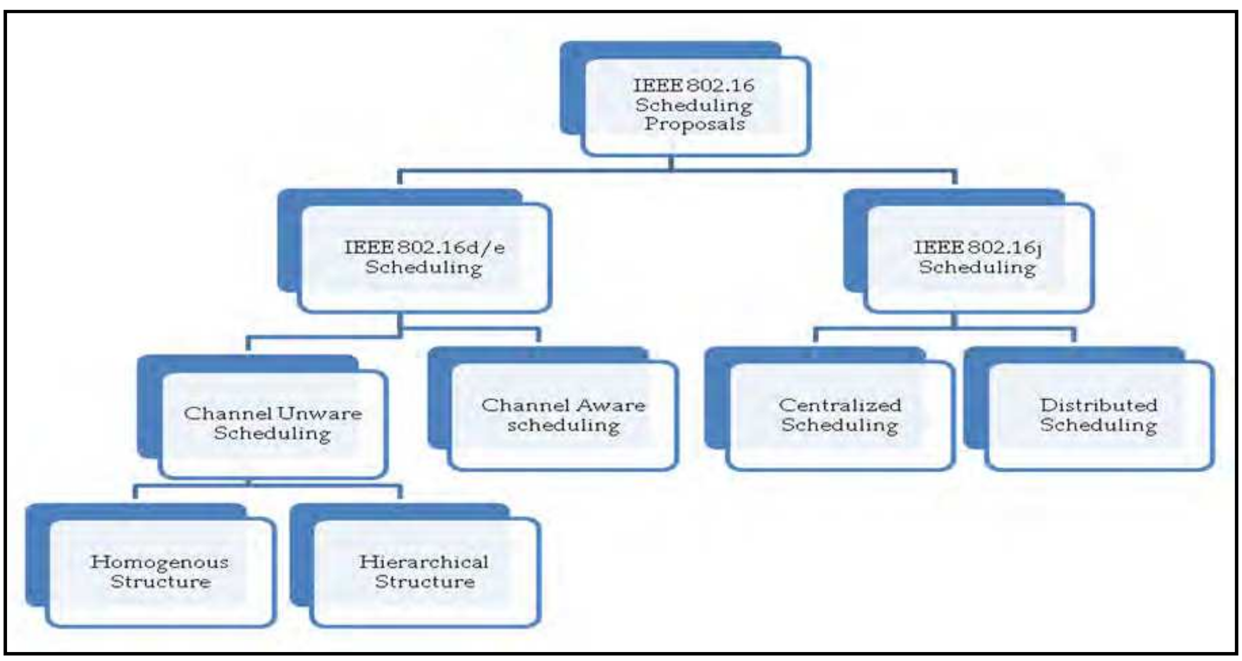


Fig. 5. Proposed classification for WiMAX Scheduling algorithms

4. IEEE 802.16e/d scheduling

4.1 Channel unaware scheduling

The algorithms belonging to this class are classical schedulers. The algorithms applied in both homogenous and hierarchical structures were originally designed for wired networks but are used in WiMAX in order to satisfy the QoS requirements. Therefore, the algorithms of this category do not consider the WiMAX channel conditions such as the channel error and loss rates.

4.1.1 Homogenous structures

Uplink homogeneous schedulers

This category of scheduling is based on simple algorithms such as Earliest Deadline First (EDF)(S.Ouled et al, 2006), Round Robin (RR), Fair Queuing (FQ), and their derivatives. A modified version of the Deficit Round Robin (DRR) is proposed in (Elmabruk et al, 2008), as a scheduling algorithm to ensure the QoS in the IEEE 802.16. The authors try to preserve the available simplicity in the original DRR design which provides O(1) complexity. The proposed scheme has one queue for both UGS and Unicast polling, and one queue for BE and a list of queues for rtPS and nrtPS. Each queue in the list represents one connection as shown in figure 6. The list is updating in each frame by adding new queues and removing the empty queues from the list. The bandwidth requirement is calculated depending on the traffic type by using the maximum sustained traffic rate r_{max} and the minimum reserved traffic rate r_{min}. Each queue in the list is related with a deficit counter variable to determine the number of the requests to be served in the round and this is incremented in every round by a fixed value called Quantum, which is computed as follow:

$$\text{Quantum} = \sum_{K=0}^{Ki} r_{min}(i, K) \qquad (1)$$

Where r_{min} is the minimum reserved traffic rate and Ki is the total connections for the i^{th} class of the service flow. An extra queue has been introduced to store a set of requests whose deadline is due to expire in the next frame.

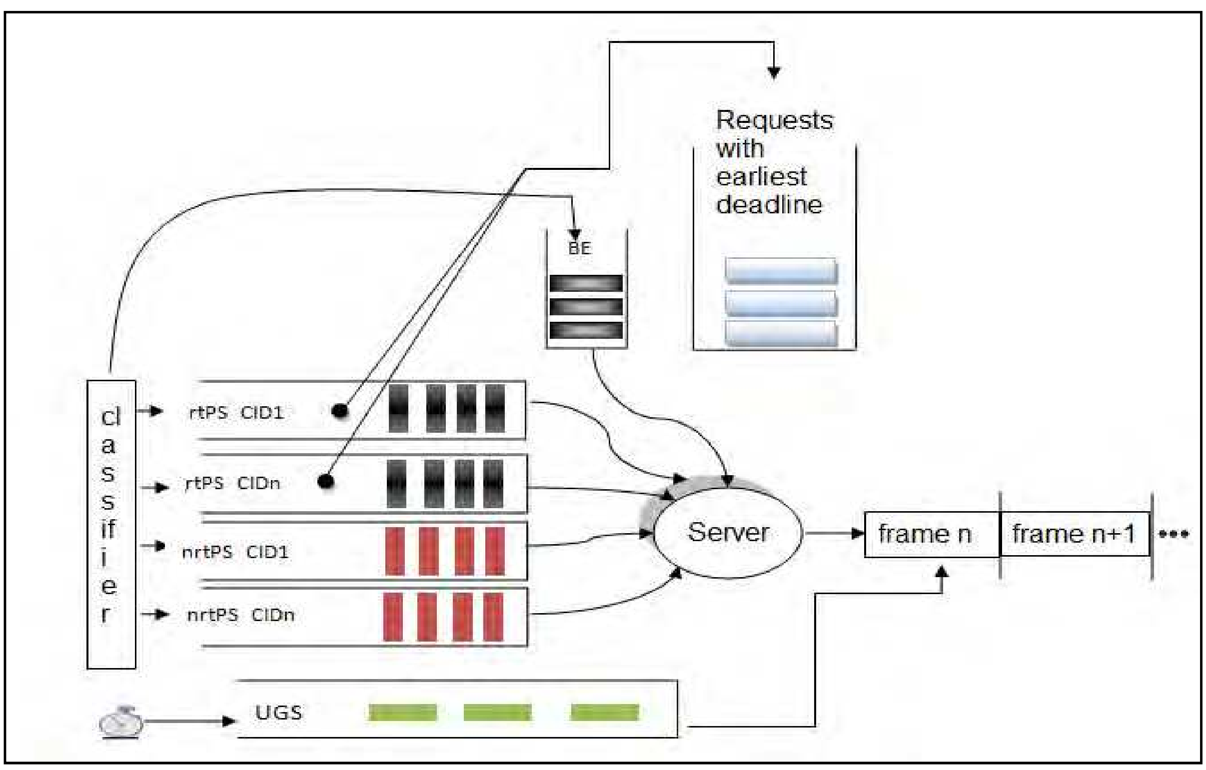


Fig. 6. Scheduler architecture proposed in (Elmabruk et al, 2008)

Every time the scheduler starts the scheduling cycle, this queue will be filled by all rtPS requests, which are expected to miss their deadline in the next frame. In the proposed scheme, it is assumed that the deadline of a request should be equal to the sum of the arrival time of the last request sent by the connection and its maximum delay requirement. In the next scheduling cycle, the scheduler will check if there are any request has been added to this extra queue. If so, the scheduler will then serve this queue after the UGS and polling queue. Once the extra queue becomes empty and there are available BW in the UL_MAP, the scheduler will continuing serving the PS list, using DRR with priority for rtPS, followed by nrtPS. For BE, the remaining bandwidth will assigned using FIFO mechanism.

In (Chirayu & Sarkar, 2009), authors propose an enhancement to the EDF principle to ensure that low priority traffic would not starved. Since the EDF tends to starve the BE traffic in presence of high number of rtPS packets. The WiMAX frame is divided into time slots, and SS are required to transmit packets in these slots, the original packets generated at the application level are fragmented to ensure that these packets fit into and can be transmitted in a time slot. When a packet is fragmented, the last fragmented packet might be of any length from 1 byte to the maximum size, which can be transmitted in a slot. If the last fragment contains lesser number of bytes than the maximum allowable fragment size, then they can stuff a part of a BE packet into this empty section. In this way, two or three such empty slots might be enough to transmit a complete BE packet to the BS. Thus, the chance that BE traffic will find an empty spaces to be transmitted is increase even there more rtPS traffic.

Downlink homogeneous schedulers

Since homogenous algorithms cannot assure the QoS guarantee for different service classes, a limited number of studies focused on this category of scheduling. RR and WRR (Cicconetti et al, 2006)(Sayenko et al, 2008) are applied in IEEE 802.16 networks in order to schedule the downlink traffic. RR algorithm allocates fairly the resources for users even they have nothing to transmit, so it may be non-conserving work scheduler and does not take into account the QoS characteristics. In WRR algorithm, the weights are assigned to adjust the throughput and latency requirements. Variants of RR such as DRR (Cicconetti et al, 2007) are applied for downlink packet scheduling in order to serve the variable size packet. The major advantage of the RR variants is their simplicity; their complexity is O (1).

In (Kim & Kang, 2005) and (Ku et al, 2006), the authors proposed a packet scheduling scheme called DTPQ (Delay Threshold-based priority Queuing) where both real time (RT) and non-real time (NRT) services are supported. The purpose of the proposed DTPQ scheduling scheme aims to maximize the number of users in the system and increasing the total service revenue. The main important parameters taken into account in this scheduling policy is the weight of both RT and NRT services denoted by ω_{RT} and ω_{NRT} respectively. The downlink packet-scheduling scheme proposed in (Kim & Kang, 2005) does not address how the delay threshold can be set while an adaptive version of DTPQ scheme is implemented in (Ku et al, 2006). In fact, the delay threshold is updated based on the variation of the weighted sum of the delay for the most urgent RT users and average data rate for RT users.

4.1.2 Hierarchical structures

Uplink hierarchical schedulers

In (Kitti & Aura, 2003), authors introduce a hierarchical structure of bandwidth allocation for IEEE 802.16 systems. Figure 7 shows a sketch of the proposed implementation UPS (Uplink Packet Scheduling). In the first level, the entire bandwidth is distributed in a strict priority manner. UGS has the highest priority, then rtPS, nrtPS, and finally BE. So inter class fairness is not achieved in presence of large number of the higher priority packets. In the second level, different mechanisms are used to control the QoS for each class of service flow.

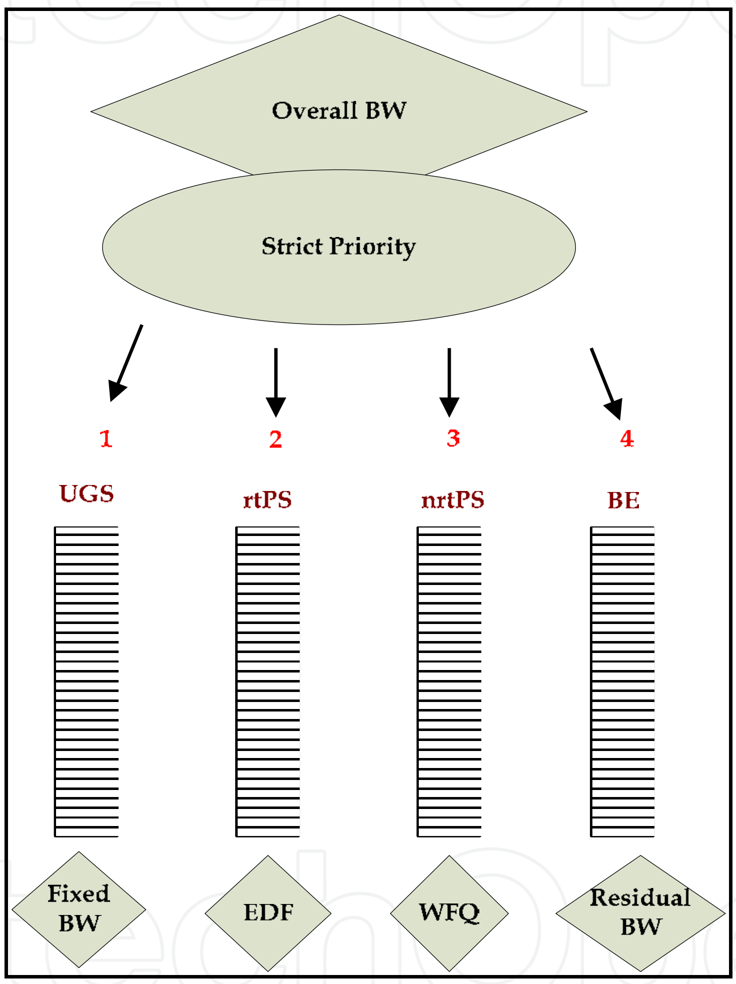


Fig. 7. Hierarchical structure proposed in (Kitti & Aura, 2003)

The uplink packet scheduler allocates fixed bandwidth to the UGS connections. Earliest deadline first (EDF) is used to schedule rtPS service flows, in which packets with the earliest deadline are scheduled first. The nrtPS service flows are scheduled using the weight fair queuing (WFQ) based on the weight of the connection. The remaining bandwidth is equally allocated to each BE connection. The UPS solution is composed of three modules: information, scheduling database and service assignment modules. Here is a brief description of the different of UPS:

- At the beginning of each time frame, the Information Module collects the queue size information from the BW-Requests received during the previous time frame. The

Information Module will process the queue size information and update the Scheduling Database Module.

- The Service Assignment Module retrieves the information from the Scheduling Database Module and generates the UL-MAP.
- BS broadcasts the UL-MAP to all SSs in the downlink subframe.
- BS's scheduler transmits packets according to the UL-MAP received from the BS.

Authors in (Tsu-Chieh et al, 2006) present an uplink packet scheduling with call admission control mechanism using the token bucket. Their proposed CAC is based on the estimation of bandwidth usage of each traffic class, while the delay requirement of rtPS flows shall be met. Each connection is controlled by token rate r_i and bucket size bi. Then, they find an appropriate token rate by analyzing Markov Chain state and according to delay requirements of connections. In their Uplink Packet Scheduling Algorithm, they adopt Earliest Deadline First (EDF) mechanism proposed in (Kitti & Aura, 2003). There is a database that records the number of packets that need to be sent during each frame of every rtPS connection. The disadvantage of this mechanism is that depends on the estimation model that is used.

In (Yanlei & Shiduan, 2005), authors propose a hierarchical packet scheduling model for WiMAX uplink by introducing the "soft-QoS" and "hard-QoS" concepts as shown in figure 8. The rtPS and nrtPS traffic are classified as soft-QoS because their bandwidth requirement varies between the minimum and maximum bandwidth available for a connection. UGS traffic is classified as hard-QoS. The model is able to distribute bandwidth between BE and other classes of traffic efficiently and guarantees fairness among the QoS-supported traffic (UGS, rtPS and nrtPS).

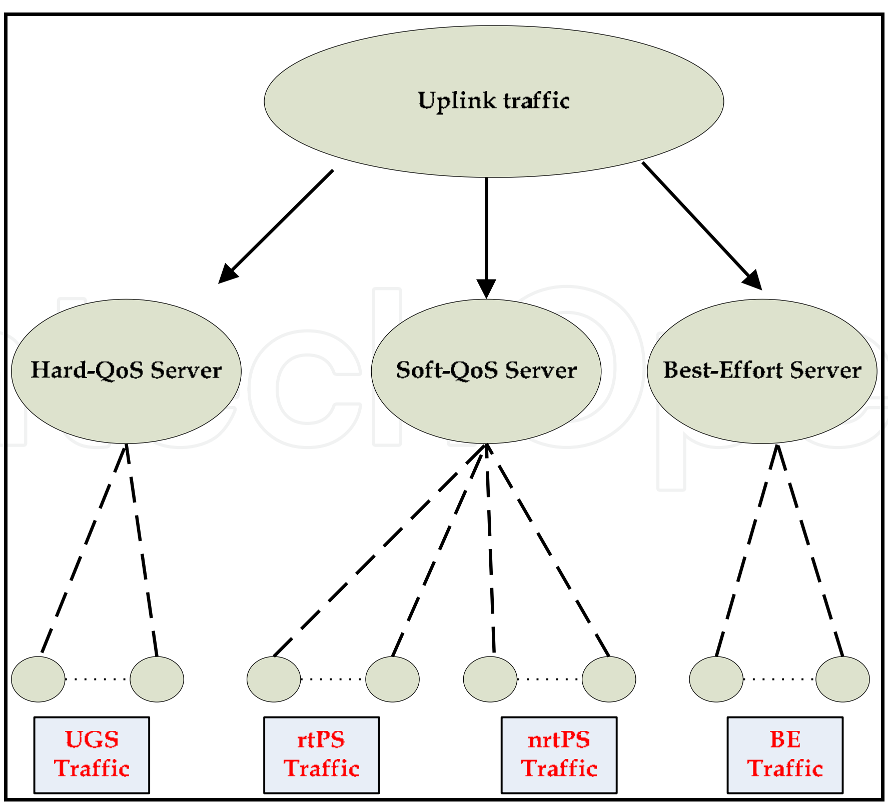


Fig. 8. Hierarchical structure as proposed in (Yanlei & Shiduan, 2005)

The packet-scheduling algorithm comprises of four parts:

1. hard-QoS server scheduling
2. soft-QoS server scheduling
3. best-effort server scheduling
4. Co-scheduling among the above three servers

The four servers implement WFQ (Weighted Fair Queuing) in their queues, for the first three servers a virtual finish time for each packet has to be calculated. The weight must be the weight of the packet and the packet having the smallest time is put at the head of the queue. The co-scheduling server calculates a virtual finish time too but here the weight should be the weight of the queue and the packet with the smallest time is served firstly.

A new distributed uplink Packet scheduling algorithm is proposed in (Sonia & Hamid, 2010). When uplink capacity cannot satisfy the required resource of connections, the traffic of one or some user terminals from user terminals in the overlapping cells are selected for transferring to the neighboring under-loaded cells. The algorithm is described as follow:

In the first step, the service assignment module, as proposed in (Kitti & Aura, 2003), calculates the uplink free capacity and resources required for each connection of a user terminal using the information saved in the scheduling database.

In the second step using the information calculated in the previous step and the traffic characteristics of the scheduling services of the user terminals, the BS checks the uplink free capacity in each time frame. If the free capacity is not enough to be allocated to necessary connections, the BS concludes that a handover is needed.

Authors in (Ridong et al, 2009) propose a utility-based dynamic bandwidth allocation algorithm in IEEE 802.16 networks to minimize the average queuing delay. The utility function is introduced as a supplementary unit, which is related to the average queuing delay of each SS node, is constructed, for QoS consideration, weight factors are introduced for different type of services. The utility function is expressed as follows:

$$U_{i,k}\left(B_{i,k}^{alloc}\right) = 1 - e^{\frac{-1}{\alpha_k \times \overline{D_{i,k}}}} \quad (2)$$

The disadvantage of the hierarchical structure is the starvation of the lower priority classes by the high priority classes.

In order to avoid this drawback, in (Chafika, 2009), authors develop an algorithm called courteous algorithm that consists of servicing the lower priority traffic without affecting the high priority traffic. Authors analyze two queues c1 and c2, which related respectively to rtPS and nrtPS classes. Packets of the c1 class have priority Pr1, while those of class c2 have priority Pr2.

Four conditions must be satisfied before applying the courteous algorithm in order to serve packet of class c1 before those of c2. These conditions are as follow:

1. Pr1>Pr2
2. $\eta_1(t) < \omega_1$
3. $\eta_2(t) > \omega_2$
4. $\tau_2 < \xi_1$

The first condition is that the priority of queue c1 is higher than that of queue c2. In the second one ω1 represents the tolerated threshold of packet loss rate for class c1 traffic, and η_1 represents the packet loss probability at time t for the class c1, which must not reach the value of ω1. The third condition relates to the probability of packet loss for class c2, which is η_2 at time t just before the application of the courteous algorithm. η_2 is the factor that determines if class c2 traffic needs more bandwidth and ω_2 represents the tolerated threshold of packet loss rate for class c2 traffic. Thus, if η_2 is greater than ω_2, then the packets of this class require to be served. In the fourth condition τ_2 is the time that required to service class c2 packets and should not exceed the tolerated waiting time ξ1 of packets of class c1. The main idea of the courteous algorithm consists in substituting service of packet of high priority with service to lower priority traffic whenever possible. This scheduling scheme is recommended when nrtPS traffic is important with respect to rtPS traffic. One more advantage of this proposal is that it improves indirectly the overall traffic since it contributes to the reduction of the packet loss rate.

Downlink hierarchical schedulers

In (Xiaojing, 2007), an Adaptive Proportional Fairness (APF) scheduling algorithm, was proposed, which is designed to extend the PF scheduling algorithm to a real time service and provides a satisfaction of various QoS requirements. The proposed APF try to differentiate the delay performance of each queue based on the Grant per Type-of- Service (GPTS) principle. The introduced priority function for queue i is defined as:

$$\mu_i(t) = \frac{r_i(t)}{R_i(t) / C_i(t) M_i(t)} \tag{3}$$

Where $r_i(t)$ is the current data rate, $R_i(t)$ denotes an exponentially smoothing average of the service rate received by SS i up to slot t. $M_i(t)$ is the minimum rate requirement, $C_i(t)$ is the number of connections of the i^{th} queue. Each queue corresponds to one QoS class, respectively. The queue having the highest value of $\mu_i(t)$ is served first. So the priority can be respectively UGS, ertPS, rtPS, nrtPS, and BE. In fact, the packets with minimum deadline or latency are measured at the highest priority level.

In (N.Wei et al, 2011), the authors proposed a QoS priority and fairness scheduling scheme for downlink traffic which guarantees the delay requirements of UGS, ertPS and rtPS service classes. The proposed mechanism is a two-level scheduling scheme that intends to maximize the BE traffic throughput. Firstly, a strict priority between service classes is adapted in the first level as follows UGS > ertPS > rtPS > nrtPS > BE. Secondly, a fixed-size data is granted periodically for UGS service class, an Adaptive Proportional Fairness (APF) scheduling is applied for both rtPS and ertPS service classes, and a Proportional Fairness (PF) scheduling is used for nrtPS and BE service classes.

A comparative study in (Y.Wang et al, 2008) is presented, compared with RR, and PF schemes, APF algorithm outperforms in service differentiation and QoS provisioning. APF is flexible to the system size in terms of the number of I accommodated users.

The priority order applied may starve some connections of lower classes. In (J.Chen et al, 2005b), a Deficit Fair Priority Queue (DFPQ) is introduced in order to reduce the problem of lower priority classes' starvation. A DFPQ is deployed in the first layer with counter to serve different types of service flows in both uplink and downlink. The counter is deceases according to the size of the packets. The scheduler moves to the next class when the counter returns to zero. Three different scheduling algorithms are used for each traffic class in the second layer. The proposed scheme is as shown in figure 9 EDF for rtPS traffics, WFQ for nrtPS class and RR for BE class. A DFPQ is better than the strict priority scheduling in order to achieve the fairness among classes.

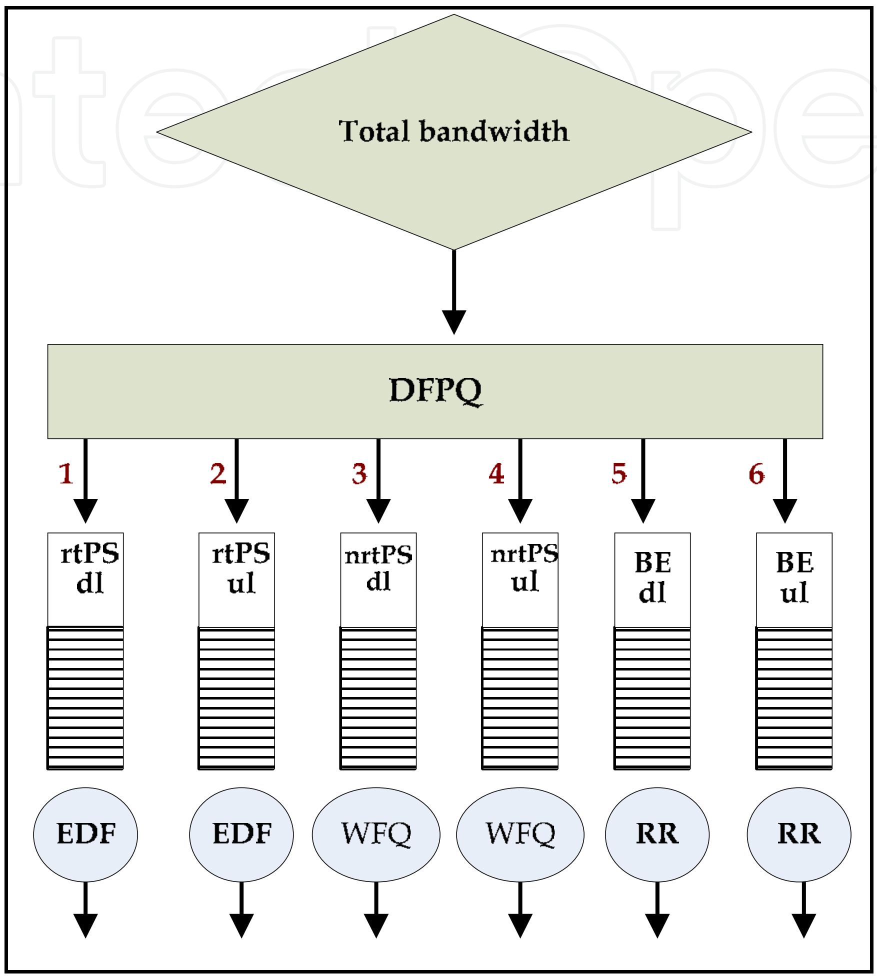


Fig. 9. Deficit Fair Priority Queue (DFPQ) as proposed in (J.Chen et al, 2005b)

4.2 Channel aware scheduling

Uplink aware schedulers

This category is also called opportunistic scheduling algorithms that is proposed for WiMAX and exploit variation in channel quality giving priority to users with better channel quality, while attempting to satisfy the QoS requirements of the multi-class traffic. A cross layer scheduling is proposed in (G.Wei et al, 2009) designed for WiMAX uplink, considering the states of queues, the channel conditions and the QoS requirements of service classes, authors propose a cross layer designed scheduling algorithm called DMIA (Dynamic MCS and Interference Aware Scheduling Algorithm) which can dynamically adapt the varying modulation and coding scheme (MCS) and the interferences in wireless channel. The objective is to maximize the total throughput, while satisfying the QoS requirement of different service classes. So, it is a constrained optimization problem.

Frequently, the cross layer algorithms formulate the scheduling problem as an optimization problem.

The DMIA proposed in (G.Wei et al, 2009) is designed to two-stage. On the first one, the dynamic bandwidth values are set for the five service classes. Therefore, the algorithm can prevent the high priority traffics from occupying too much bandwidth resources, and adjust the amount of scheduling data according to the varying MCS. On the second stage, different connections belong to the same service class will be scheduled according to the priority functions.

In (Liu et al, 2005), the authors propose a priority- based scheduler at the MAC layer for multiple connections with divers QoS requirements, where each connection employs adaptive modulation and coding (AMC) scheme at the physical layer. The authors define a priority function that integrates in its formulation the delay of HOL packet and the minimum required bandwidth. Each non-UGS connection admitted in the system is assigned with a priority, which is updated dynamically based on the channel quality and on its service class. The number of time slots allocated per frame to UGS connections is fixed. The proposed scheduler enjoys flexibility since it does not depend on any specific traffic or channel model. Besides, in (Pratik, 2007) authors have chosen to evaluate the performances of the proposed cross layer in (Liu et al, 2005), their evaluation indicates high frame utilization as it indicates poor performance with respect to average throughput, average delay and fairness.

A Cross-Layer Scheduling Algorithm based on Genetic Algorithm (CLSAGA) under the Network Utility Maximization (NUM) concepts is proposed in (Jianfeng et al, 2009). Adaptive modulation and coding (AMC) scheme and QoS category index of each service flow jointly decide the weights of utility functions to calculate the scheduling scheme of MAC layer. The genetic algorithm can be used to solve optimization problems. The cross layer diagram is shown in figure 10.

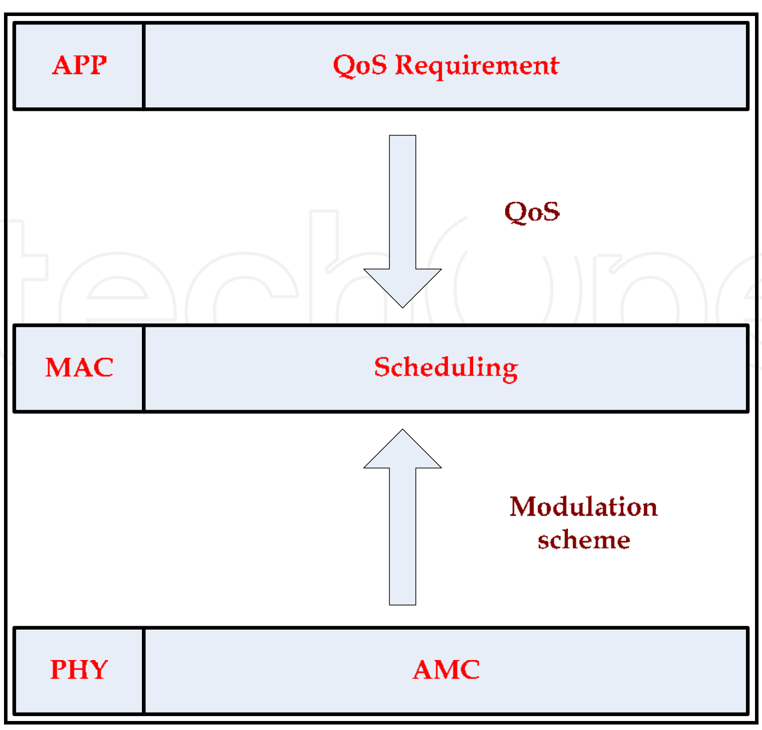


Fig. 10. Cross layer diagram as proposed in (Jianfeng et al, 2009)

Downlink aware schedulers

In (Hongfei et al, 2009), the authors proposed a practical cross-layer framework for downlink scheduling with multimedia traffic called CMA (Connection-oriented Multistate Adaptation) illustrated in figure 11. A multisession MBS scheduling in multicast/Broadcast (MC/BC)-based WiMAX is taken into account in the proposed scheme. The authors adopt the service-oriented design on per-service-flow carrying multisession MBS. The framework performs simultaneous adaptations across protocol stacks on source coding, queue prioritization, flow queuing, and scheduling. CMA achieves the lowest variance value with the fastest convergence curve and lowest max-min variations, which mean that it can provide SSs with better throughput equality in a short time.

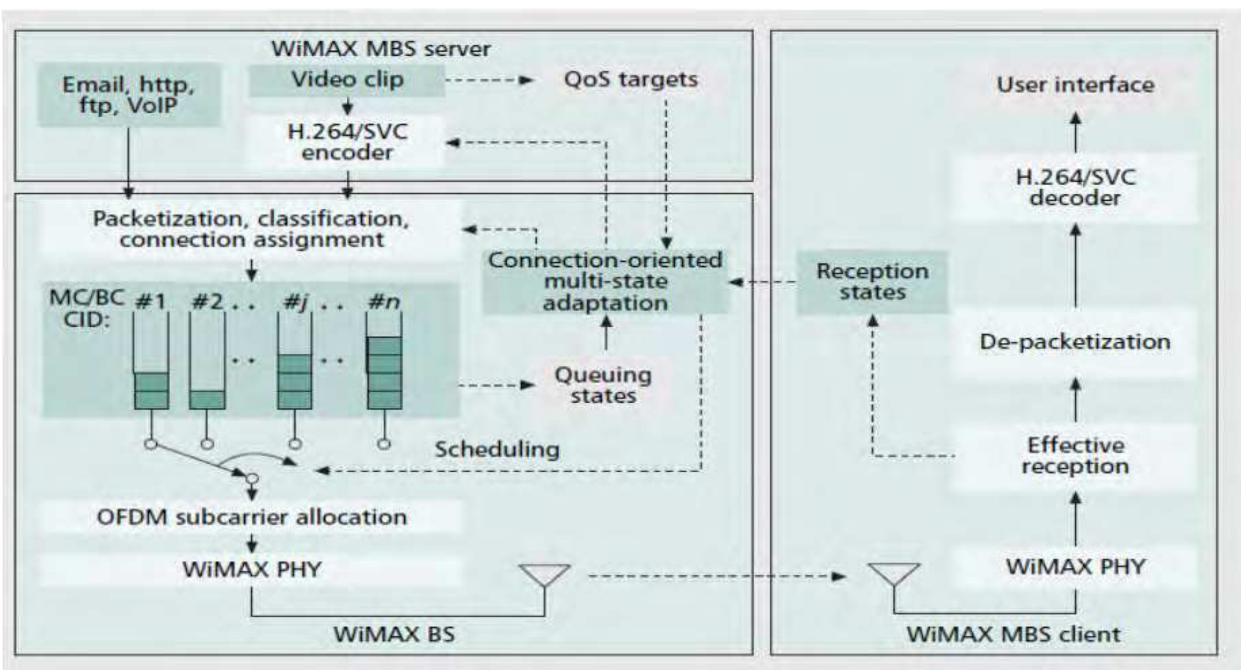


Fig. 11. CMA scheduling framework proposed in (Hongfei et al, 2009)

In (Vishal et al, 2011), the authors proposed a resource allocation mechanism for downlink OFDMA, which aims to maximize the total throughput with lesser complexity while maintaining rate proportionally between users. The BS allocates sub carries to existing users and the number of bits per OFDMA symbol from each user to be transmitted on each sub carries.

The main steps of the proposed mechanism are described as follows:

- Calculate number of subcarriers assigned to each user.
- Assign subcarriers to each user to achieve proportionality.
- Assign total power P_k to each user maintaining proportionality.
- Assign $P_{k,n}$ for each users subcarriers subjected to P_k constraints. Where $P_{k,n}$ is the power assigned per subcarrier per user.

5. IEEE 802.16j scheduling

Unlike in single hop networks, in a Mobile Multihop Relays (MMR) system (Standard, 2008), service scheduling more complicated. Because the BS need to discover if all the RSs (relay stations) in the path to the SS have sufficient resources to support the QoS request. The discovery procedure begins with the BS sending a DSA request message to its subordinate RS. Then the RS sends its own DSA request message to its subordinates RSs and so on until the access RS is reached.

There are two different options of scheduling in MMR networks: centralized and distributed. In the first option, the BS performs the scheduling of all nodes, while in the second option; the relay stations have certain autonomy and can make scheduling decisions for nodes in communications.

An IEEE 802.16j frame structure is divided into access and relay zones, as well as the uplink sub frame that is divided into access zone and relay zone. The IEEE 802.16j standard defines two kinds of relay:

1. Transparent relays: Access zone is used by SS to transmit on access links to the BS and RSs. The relay zone is used by RSs to transmit to their coordinates RSs or BS. This kind of relay operates only in centralized scheduling mode within the topology of maximum two hops.
2. Nontransparent relays: This mode introduces the multihop scenario. There are two ways of transmitting and receiving the frame. The first one is to include multiple relay zones in a frame and relays can alternately transmit and receive in the different zones. The second one is to group frames together into a multi-frame and coordinate a repeating pattern in which relays are receiving or transmitting in each relay zone.

There are three cases of SS/BS communication:

i. The SS is connected to BS directly.
ii. The SS is connected to the BS via a transparent relay.
iii. The SS is connected to the BS via one or more nontransparent relays.

5.1 Distributed scheduling

Uplink relay schedulers

In (Debalina et al, 2010), authors propose a heuristic algorithm, OFDMA Relay Scheduler (ORS) algorithm, for IEEE 802.16j networks. The ORS algorithm is used to schedule traffic for every SS/RS in each scheduling period. A scheduling period consists of an integral number of frames. The ORS scheduler works for all three cases of SS/BS communication and it consists of two main parts:

- Frame division and Bandwidth Estimation:

The frame relay zone is divided into even and relay zone to maintain the half-duplex nature of the node. So, nodes are labeled alternately even or odd. Even nodes transmit in even relay zone and odd nodes transmit in odd relay zone. The BS is assigned an even label. Thus, the children of the BS are labeled odd.

For the Bandwidth Estimation, if the BS obtains information about the CINR (Carrier to Interference-plus-Noise-) and RSSI (Received Single Strength Indication) values, it can determine the data rate used by the sub channel. Therefore, if the BS does not know about the CINR and RSSI values, then the ORS algorithm compute the lower bound of the network capacity by assuming all the slots available for data transmission are modulated at the most robust and least rate.

- Slot Scheduling:

The ORS heuristic schedules slots for a particular service class to all the nodes in a zone before considering the next zone. The proper zone where the slots for a particular node will be allocated is based on whether the child is a MS or RS and the label of an RS. The node is then allocated slots based on the best available sub channels, which are picked for scheduling the link based on their CINR and RSSI values.

The proposed ORS in (Debalina et al, 2010) addresses adaptive zone boundary computation, determination of schedule for prioritized traffic based on traffic demand while incorporating frequency selectivity within a zone and adapting to changing link conditions in IEEE 802.16j networks.

Downlink relay schedulers

In (Yao et al, 2007), a factor-graph-based low-complexity distributed scheduling algorithm in the downlink direction is proposed. The proposed algorithm manages excellent performance by exchanging weighted soft-information between neighboring network nodes to obtain a series of valid downlink transmission schedules that lead to high average values and low standard deviations in packet throughputs.

A factor graph consisting of agent nodes, variable nodes, and edges is a graphic representation for a group of mutually interactive local constraint rules. Soft-information indicates the probability that each network link will be activated at each packet slot. The proposed scheme consists of three main parts are described as follows:

- **Factor Graph Modeling and Sum-Product Algorithm:** the factor graph model is constructed in order to model the example network scenario and to specify all the local constraint rules enforced by each agent node. The local rules specified by the BS denoted by *B,* two relays *R*1 and *R*2, and four MSs *M*1, *M*2, *M*3 and *M*4.
- **Calculation and Transportation of Soft-Information**: Four iterative steps are implemented in this part. In step1, an initialization of soft-Information is done to indicate the transfer from a node to another one. The second step processes the passage of the soft-Information from a variable node to one of the agent node. The third step assigns weights to the soft-information of a local transmission pattern according to the network traffic condition. Finally, in the fourth step, the stop criterion is set. The proposed algorithm sets the maximum number of iterations at 10. If the number of iterations exceeds this number, the algorithm will stop and the procedure will restart from the initialization step.
- **A Feasible-Weighting Scheme:** A heuristic and feasible weighting scheme is defined. Weight is assigned for each local transmission pattern differently in order to increase

resource utility. Thus, the information required for the weighting scheme must be locally achievable.

5.2 Centralized scheduling

Uplink relay schedulers

A traffic adaptive uplink-scheduling algorithm for relay station is proposed in (Ohym & Dong, 2007). It focuses on the system transparency in IEEE 802.16j. The aims of this algorithm are to minimize the end-to-end delay and signaling overhead and to avoid the resource waste. Authors in (Ohym & Dong, 2007) consider two main strategies: one is elaborated for the real time service flows, and the other is elaborated for the non real time service flows. The first strategy has to allocate resources for RS based on the bandwidth request information of the Mobile Station (MS) is defined as MS-REQ since it use the bandwidth request information of the MS. BS allocates bandwidth for uplink data transmission at each frame based on the bandwidth request information of only MSs without any bandwidth request of the RS. The MS-REQ is as follow:

While any service flow exists

1. BS allocates bandwidth for MS and RS at a time
2. MS transmits data to RS

3- a) On receiving the data from MS, RS transmits the received data by using pre-allocated resource

3- b) If the data from MS is broken, RS transmits nothing

End

Allocating bandwidth for relay station in advance may generate resource waste. If the data is broken between the MS and the RS, the pre-allocated bandwidth is not used and the bandwidth efficiency cannot be maximized only with MS-REQ method. Thus, this scheduling strategy is suitable for the case of light traffic load.

The second strategy is defined as TR-QUE. It has to allocate resources for RS based on the direct bandwidth request of RS. The relay station queues the received data from mobile stations according to the existing scheduling classes: UGS, rtPS, ertPS, nrtPS, and BE. Then, the RS regards each queue as each service flow. The TR-QUE is detailed as follow:

While any service flow exists

1. BS firstly allocates bandwidth only for MS
2. MS transmits data and RS receives and queues the data
3. RS requests bandwidth for successful data
4. BS allocates bandwidth for RS and RS transmits the data

End

It is the optimized scheduling solution for RS in term of bandwidth efficiency. This scheduling strategy is suitable for the case of heavy traffic load. Since The RS does not waste resource even if some part of data-packets from the mobile stations to the relay station are broken due to poor channel. However, the delay performance cannot optimize only by this strategy. In order to optimize both the delay and bandwidth requirements, authors propose a hybrid method Hyb-REQ that uses MS-REQ method for real time traffic and TR-QUE method for non-real time traffic, respectively. The Hyb-REQ algorithm is defined as follow:

While any service flow exists

If non-real time service flow

1. BS firstly allocates bandwidth only for MS
2. MS transmits data and RS receives and queues the data
3. RS requests bandwidth for successful data
4. BS allocates bandwidth for RS and RS transmits the data

If real time traffic service flow

5. BS allocates bandwidth for MS and RS at a time
6. MS transmits data to RS

7- a. In case of success, RS transmits the received data
7- b. In case of error, RS transmits the queued data in step 2

End

When some real time service packets are broken between the MS and the RS, Hyb-REQ transmits some part of non-real time data packets queued at the RS without bandwidth request by using the pre-allocated bandwidth, which was supposed to be wasted. The Hyb-REQ scheduling improves the delay requirement for the real time service traffic using and maximizes the throughput for the non real time service. So the proposal algorithm tends to satisfy the QoS dependent on the traffic.

Downlink relay schedulers

In (Gui, 2008), two relay-assisted scheduling schemes are defined, in which the RS assists the BS in its scheduling decision and therefore it is possible for the BS to exploit CSI (Channel State Information) on the access links without those of the relay links from all the users directly. Authors consider a set of K mobile users, uniformly distributed in a cell, served by a single base station with M relay stations, in which each mobile device intends to receive its NRT data from the BS, possibly by multi-hop routing. Each user rightly predicts its own downlink channel state information and feedback information, combined with the information of the quality of service (e.g. throughput and delay) that each user has achieved so far, is used to calculate the priorities by certain scheduling algorithm at the BS side. For each time slot, either a mobile terminal or a relay terminal with the highest priority is selected by BS for the transmission of the data packets. Figure 12 describes the packet scheduler structure proposed.

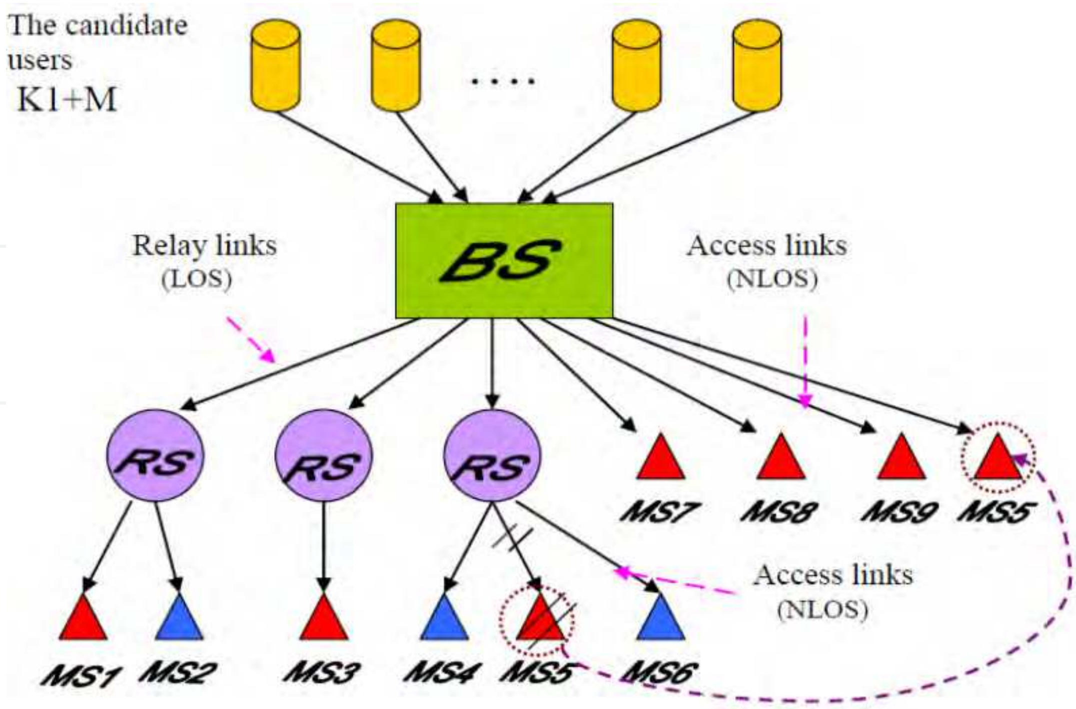


Fig. 12. Packet scheduler structure proposed in (Gui, 2008)

6. Comparative and synthesis study

The Table 4, as shown below presents a comparative analysis of the QoS Scheduling Algorithms in PMP mode.

Category	Traffic	Scheduling Proposal	Strength	Limitation	QoS aspects
Homo-genous	Uplink	(Elmabruk et al, 2008)	Simple	Unsuitable for uplink traffic	Attempt to satisfy all classes
		(Chirayu & Sarkar, 2009)	Does not starve the BE traffics	Introduce overheads	Throughput for NRT and delay for RT classes
	Downlink	(Cicconetti et al, 2006) (Sayenko et al, 2008)	Enhance the QoS satisfaction	Does not consider the channel behavior	2 types of class (rtPS, BE)
		(Kim & Kang, 2005)	Maximize the number of SS and increase the total revenue	Does not address the delay setting	Delay for RT classes and throughput for NRT
		(Ku et al, 2006)	Maximize the number of SS and increase the total revenue and The delay threshold is updated	Unstable	Maximize throughput while maintaining delay

Category	Traffic	Scheduling Proposal	Strength	Limitation	QoS aspects
Hiera-rchical	Uplink	(kitti & Aura, 2003)	Satisfy the major QoS requirements	Complex and unfair	Delay for RT traffics and throughput for NRT traffics
		(Tzu-Chieh, 2006)	Satisfy the major QoS requirements	Complex and need estimation model	Satisfy delay requirements
		(Yanlei & Shiduan, 2005)	Satisfy the main QoS requirements	Complex	3 types of service (UGS, rtPS, nrtPS)
		(Sonia & Hamid, 2010)	QoS guarantee	Complex and handover process	Attempt to serve all types of connections
		(Ridong et al, 2009)	Minimize the average queuing delay	Unfair	Delay requirements for RT classes
		(Chafika, 2009)	Fair and satisfy the QoS requirements	Complex and need mathematical model	Serve the lower priority traffic
	Downlink	(Xiaojing, 2007)	Performs throughput, fairness, and frame utilization	low average delay	All types of service
		(N.Wei et al, 2011)	increase the network throughput and lower delay	Does not support the radio channel	All types of traffic
		(J.Chen et al, 2005b)	Provides more fairness to the system	Complex implementation	All types of traffic
Aware schedulers	Uplink	(G.Wei et al, 2009)	Address the channel state condition and try to satisfy the QoS requirements	Complex	Maximize the total system throughput
		(Liu et al, 2005)	Use the AMC scheme and try to satisfy the QoS constraints	Complex	Respect to average throughput and average delay
		(Jianfeng et al, 2009)	Genetic algorithm implementation	Complex	Balances priorities between mobile stations
	Downlink	(Hongfei et al, 2009)	Viable end-to-end architecture	Complex	Delay, throughput and fairness
		(Vishal et al, 2011)	Low complexity adaptive resource allocation	Starve the lower priority traffic	Maximize total throughput

Table 4. IEEE 802.16d/e proposed methods comparison based on the proposed classification

The Table 5, as shown below presents a comparative analysis of QoS scheduling algorithms, which are dedicated to support the relay mode.

Category	Traffic	Scheduling Proposal	Strength	Limitation	QoS considered
Distributed	Uplink	(Debalina et al, 2010)	Adaptive computation to the channel conditions	complex	All types of service
	Downlink	(Yao et al, 2007)	Increase data packet throughput And increase resource utility, avoid collision	Complex Does not address the delay constraint	-
Centralized	Uplink	(Ohyun & Dong, 2007)	Enhancement of delay and bandwidth requirements	Complex	All types of service
	Downlink	(Gui, 2008)	overhead is avoided	Complex	NRT services

Table 5. IEEE 802.16j proposed methods comparison based on the proposed classification

7. Conclusion

In this chapter, we have provided an extensive survey of recent WiMAX proposals that provide and enhance QoS. All the relevant QoS functionality's such as bandwidth allocation, scheduling, admission control, physical modes and duplexing for WiMAX are deeply discussed. Call Admission Control (CAC) is an important QoS component in WiMAX networks as it has a strong relationship with QoS parameters such as delay, dropping probabilities, jitter and scalability. Therefore, we present a classification and a description of CAC algorithms proposed in the literature for PMP mode. We describe, classify, and compare CAC proposals for PMP mode. Although many CAC scheme has be introduced in the literature, there is stillroom for improvement CAC mechanism.

The QoS platform designers need to be familiar with WiMAX characteristics. So in this chapter, we have present cross-layer designs of WiMAX/802.16 networks. A number of physical and access layer parameters are jointly controlled in synergy with application layer to provide QoS requirements. Most important QoS key concepts are identified. Relations and interactions between QoS functional elements are discussed and analyzed with cross layer approach consideration.

Moreover, scheduling is a main component of the MAC layer that assures QoS to various service classes. Scheduling algorithms implemented at the BS has to deal with both uplink and downlink traffics. An understanding classification of the uplink and downlink scheduling in the IEEE 802.16 networks is described in details. We present a survey of some scheduling research in literature for WiMAX fixe, mobile, and relay. In order to give a comparative study between the proposals mechanisms, we draw two summary tables showing the strength, the limitation and QoS observed aspect of each scheduling method proposed for fixed, mobile and relay WiMAX network. We have discussed the approaches and key concepts of different scheduling algorithms which can be useful guide for further research in this field.

As the scheduling in WiMAX wireless network is a challenging topic, future works should include advanced investigations on scheduling algorithms under different CAC schemes and bandwidth allocation mechanisms.

Furthermore, we intend to evaluate the behavior and the efficiency of some scheduling and CAC modules for the mobile and the relay WiMAX networks under full saturation condition and to provide a mathematical analysis combined with extensive simulations.

8. References

A. Sayenko, O. Alanen, and T. Hamaainen. (2008). Scheduling solution for the IEEE 802.16 base station, Int*. J. Computer and Telecommunications Networking*, vol. 52, pp. 96-115, Jan. 2008.

C. Cicconetti, L. Lenzini, placeE. Mingozzi, and C. Eklund. (2006). Quality of service support in IEEE 802.16 networks, *IEEE Network*, vol. 20, pp. 50-55

C.Cocconetti, A.Erta, L.Lenzini and E.Mingozzi. (2007). Performance Evaluation of IEEE802.16 MAC for QoS Support, *IEEE Transactions on Mobile Computing*, vol.6,no1,pp.26-38

C.-M. Chou, B.-J. Chang, and Yung-Fa Huang. (2006). Dynamic Polling Access Control for High Density Subscribers in Wireless WiMAX Networks, *Proceedings of the 2006 Taiwan Area Network Conference (TANET'06)*, Hualien, Taiwan, Nov. 2006

Carlos Valencia, (2009). Scheduling Alternatives for Mobile WiMAX End-to-End Simulations and Analysis, Master of Applied Science in Technology Innovation Management, Carleton University Ottawa, Canada, K1S 5B6

Chafika TATA, (2009). *Algorithme de Courtoisie : optimisation de la performance dans les réseaux WiMAX fixes*, Master thesis 2009.

Chirayu Nagaraju and Mahasweta Sarkar, (2009). A Packet Scheduling To Enhance Quality of Service in IEEE 802.16, *Proceedings of the World Congress on Engineering and Computer Science* Vol IWCECS 2009.

D.H. Kim and C. G. Kang. (2005). Delay Threshold-based Priority Queuing Packet Scheduling for Integrated Services in Mobile Broadband Wireless Access System, *Proceedings of the IEEE Int. Conf. High Performance Computing and Communications*, Kemer-Antalya, Turkey, pp.305-314.

Debalina Ghosh, Ashima Gupta, Prasant Mohapatra. (2009). Adaptive Scheduling of Prioritized Traffic in IEEE 802.16j Wireless Networks, *Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, (WIMOB'09*), pp 307 - 313

Elmabruk. Laias, Irfan Awan, et Pauline.ML.Chan. (2008). An Integrated Uplink scheduler In IEEE 802.16, *Proceedings of the second UKSIM European Symposuim on Computer Modeling and Simulation*

F.Rango, A.Malfitano and S.Marano, (2011). GCAD: A Novel Call Admission Control Algorithm in IEEE 802.16 based Wireless Mesh Networks, JOURNAL OF NETWORKS, VOL. 6, NO. 4, p 595-606

Gui Fang. (2008). Performance Evaluation of WiMAX System with Relay-assisted Scheduling, Master Thesis 2008

H. LABIOD, H. AFIFI, WI-FI, (2007). *ZIGBEE AND WIMAX*, 327 pages, 978-1-4020-5396-2 (HB) ISBN 978-1-4020-5397-9 (e-book), Published by Springer 2007.

H. Zhu and K. Lu. (2006). Above Packet Level Admission Control and Bandwidth Allocation for IEEE 802.16 Wireless MAN, Simulation Modeling Practices and Theory, 24 November 2006

H.Wang, W. Li, and D. P. Agrawal, (2005). Dynamic admission control and QoS for 802.16 wireless MAN, in Wireless Telecommun. Symp. pp. 60–66

Hongfei Du, Jiangchuan Liu, and Jie Liang. (2009). Downlink Scheduling For Multimedia Multicast/Broadcast Over Mobile WiMAX: Connection-Oriented Multistate Adaptation, *IEEE Wireless Communications*, August 2009

IEEE Draft Standard P802.16j/D5, (2008). Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems -Multihop Relay Specification," May 2008.

IEEE Standard for Local and metropolitan area networks. (2006). Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems" Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, 28 February 2006.

IEEE Standard. (2004). Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems", 3 Park Avenue, New York, NY 10016-5997, USA, IEEE Computer Society and the IEEE Microwave Theory and Techniques Society Sponsored by the LAN/MAN Standards Committee, Print: SH95246, PDF: SS95246

J. Chen, W. Jiao and H. Wang. (2005). A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode, *Proceedings of IEEE International Conference on Communications*, Seoul, Korea, vol. 5, pp.3422-3426.

J. M. Ku, S. K. Kim, S. H. Kim, S. Shin, J. H. Kim, and C. G. Kang. (2006). Adaptive delay threshold-based priority queuing scheme for packet scheduling in mobile broadband wireless access system, *Proceedings of the IEEE Wireless Communication and Networking Conf.*, Las Vegas, NV, vol. 2, pp. 1142-1147.

J.Chen, Wenhua Jiao and Qian Guo. (2005). An integrated QoS control architecture for IEEE 802.16 broadband wireless access systems, *Proceedings of the Global Telecommunications Conference*, St. Louis, MO, pp. 6-11, ISBN: 0-7803-9414-3

Jeffrey G. Andrews, (2007). *Fundamentals of WiMAX Understanding Broadband Wireless Networking*, ISBN 0-13-222552-2,478 pages, Prentice Hall Communications Engineering and Emerging Technologies Series, text printed in the United States in Westford, Massachusetts. First printing, February 2007

Jianfeng Song, Jiandong Li, Changle Li, (2009). A Crosss-Layer Scheduling Algorithm based on Genetic Algorithm, *Proceeding of the seventh Annual Communication Networks and services Research Conference*.

K.Yu, X.Wang, S.Sun, L. Zhang, X.Wu, (2009). A Statistical Connection Admission Control Mechanism for Multiservice IEEE 802.16 Network, Proceeding of 69th IEEE International conference on Vehicular Technology Conference, VTC Spring 2009

Kalikivayi Suresh, Iti Saha Misra and Kalpana Saha. (2008). Bandwidth and delay guaranteed call admission control for QoS provisioning in IEEE 802.16e Mobile WiMAX, *Proceedings of the GLOBECOM*, 2008

L. Wang, F. Liu, Y. Ji, and N. Ruangchaijatupon. (2007). Admission control for non-preprovisioned service flow in wireless metropolitan area networks, *Proceedings of the Fourth European Conference of Universal Multiservice Netw. ECUMN '07*, 2007, pp. 243–249

Loutfi Nuaymi, (2007). *WiMAX: Technology for Broadband Wireless Access*, John Wiley & Sons 2007 (310 pages), ISBN:9780470028087

M. Castrucci, F. Delli Priscoli, C. Buccella, V. Puccio, I. Marchetti. (2008). Connection Admission Control in WiMAX networks, *Proceedings of the 17th ICT Mobile and Wireless Communications Summit (ICTMobileSummit)*, Stockholm Sweden, June 2008

Ohyun Jo and Dong-Ho Cho. (2007). Traffic adaptive uplink scheduling scheme for relay station in IEEE 802.16 based multi-hop system", *Proceedings of the 66th Conference on Vehicular Technology, (VTC-2007), pp 1679 - 1683*

Pratik Dhrona. (2007). *A Performance Study of Uplink Scheduling Algorithms in Point to Multipoint WiMAX Networks*, Master thesis, December 2007.

Q.Liu, X.Wang and G.Giannakis. (2005). Cross-layer scheduler design with QoS support for wireless access networks, *Proceedings of International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*

Ridong Fei, Kun Yang, Shumao Ou, et Wenyoung Wang. (2009). A Utility-based Dynamic Bandwidth Allocation Algorithm in IEEE 802.16 NETWORKS to Minimize Average Queuing Delay, *Proceedings of the International Conference on Communication and Mobile Computation 2009.*

S. Chandra and A. Sahoo, (2007). An efficient call admission control for IEEE 802.16 networks, *Proceedings of 15th IEEE LAN/MAN Workshop, LANMAN 2007*, pp. 188–193.

S. Ould Cheikh El Mehdi, W. Fawaz, Ken Chen. (2006). Une nouvelle approche pour la gestion de flux temps réels basée sur l'algorithme EDF, *Colloque Francophone sur l'ingénierie des Protocoles CFIP'06*

Shafaq B. Chaudhry, Ratan K. Guha.(2007). Adaptive connection admission control and packet scheduling for QoS provisioning in mobile wiMAX, *Proceedings of IEEE International Conference on Signal Processing and Communications (ICSPC 2007*), Dubai, United Arab Emirates, 2007

Shepard, Steven, (2007). *WiMAX Crash Course*, ISBN: 0072263075, EAN: 9780072263077, pages 339 edition McGraw-Hill, 1/6/2006

Shida Luo, Zisu Li, (2008). An Effective Admission Control for IEEE802.16 WMN, *Proceedings of the second International Symposium on Intelligent Information Technology Application,* 2008

Sonia Nazari and Hamid Beigy. (2010). A New Distributed Uplink Packet Scheduling Algorithm In WiMAX Newtorks, *Proceedings of the Second International Conference Future Computer and Communication (ICFCC'10)*

Tzu-Chieh Tsai, Chi-Hong and Chuang-Yin Wang. (2006). CAC and Packet Scheduling Using Token Bucket for IEEE 802.16 Networks", *Journal of communications*, vol 1, NO 2

W. Kitti, and G. Aura. (2003). Packet Scheduling for QoS Support in IEEE 802.16 Broadband Wireless Access Systems, *International Journal on Communication Systems*. Vol.16, Issue1, pp: 81-96

Wankhede Vishal A, Jha Rakesh, Upena Dalal, (2011). Resource Allocation for Downlink OFDMA in WiMAX Systems, *Proceedings of the International Conference on Communication, Computing and Security,* ICCCS'11, India, February 2011

Wei Gan, Jiqing Xian, Xianzhong Xie, et Juan Ran. (2009). A Cross-layer Designed Scheduling Algorithm for WiMAX Uplink, *Proceedings of the Ninth International Conference on Electronic Measurement & Instruments ICEMI'2009*

Wei Nie, Houjun Wang and Jong Hyuk Park. (2011). Packet Scheduling with QoS and Fairness for Downlink Traffic in WiMAX Networks, *Journal of Information Processing Systems*, Vol.7, No.2, June 2011 DOI:10.3745/JIPS.2011.7. 2.261

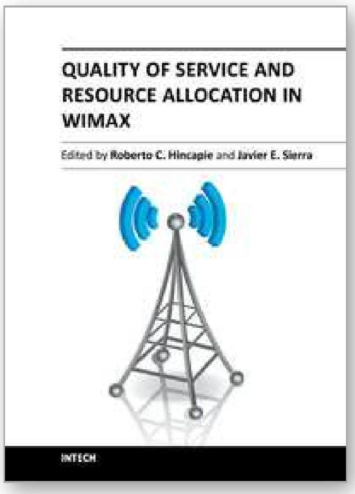
Xiaojing Meng, (2007). An Efficient Scheduling for Diverse QoS Requirements in WiMAX, Master thesis 2007.

Y. Wang, S. Chan, M. Zukerman, and R.J. Harris. (2008). Priority-Based fair Scheduling for Multimedia WiMAX Uplink Traffic, *Proceedings of the IEEE International Conference on Communications*, Beijing, China, pp. 301-305

Y.Zhang Hsiao-Hwa Chen, (2008). *MOBILE WiMAX Toward Broadband Wireless Metropolitan Area Networks*, ISBN 978-0-8493-2624-0, edition Auerbach, 2008

Yanlei Shang et Shiduan Cheng, (2005). An Enhanced Packet Scheduling Algorithm for QoS Support in IEEE 802.16 Wireless Network, *Proceedings of the third International Conference on Networking and Mobile Computing (ICCNMC'05)*, China, pp.652-661.

Yao-Nan Lee, Jung-Chieh Chen, Yeong-Cheng Wang, and Jiunn-Tsair Chen. (2007). A Novel Distributed Scheduling Algorithm for Downlink Relay Networks", *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, VOL. 6, NO. 6



Quality of Service and Resource Allocation in WiMAX
Edited by Dr. Roberto Hincapie

ISBN 978-953-307-956-1
Hard cover, 376 pages
Publisher InTech
Published online 03, February, 2012
Published in print edition February, 2012

This book has been prepared to present state of the art on WiMAX Technology. It has been constructed with the support of many researchers around the world, working on resource allocation, quality of service and WiMAX applications. Such many different works on WiMAX, show the great worldwide importance of WiMAX as a wireless broadband access technology. This book is intended for readers interested in resource allocation and quality of service in wireless environments, which is known to be a complex problem. All chapters include both theoretical and technical information, which provides an in depth review of the most recent advances in the field for engineers and researchers, and other readers interested in WiMAX.