

**Efficient QoS-based Uplink Scheduling in IEEE
802.16e Mobile Stations**

Efficient QoS-based Uplink Scheduling in IEEE 802.16e Mobile Stations

Abstract— The IEEE 802.16e-2005 standard for broadband wireless access provides a rich quality of service support framework. The standard defines five classes of service. In this paper, we provide overview of a new efficient scheme for mobile station uplink scheduling. The scheme is a credit-based algorithm based on deficit round robin. The scheduling is based on three rounds, with each round satisfying specific constraints of the scheduling problem. We provide simulation results that show the excellent performance of the scheme even under extreme overload conditions. The results show the perfect isolation of service classes and the stability of the scheme. The algorithm is implemented on an ARM9 core running at a 192 MHz clock. We show that for a frame duration of 5 msec the proposed solution is capable of sustaining high-bit-rate traffic upto 3 Mbps with five connections each from a different class of the 802.16e scheduling types.

Index Terms— IEEE 802.16e, uplink scheduling, quality of service, deficit-round robin..

1. Introduction

High-speed mobile access to Internet services has been driving the cellular industry to look for new technologies beyond 2.5 and 3G, such as, HSPA, EVDO, and LTE. Moreover, while WiFi networks provide the ubiquity and high-data rates, they fail to provide true mobility and extended range. In response, the IEEE 802 group worked on extending the IEEE 802.16-2004 (previously known as 802.16d) [1]. The result was the IEEE 802.16e-2005 standard [2-3] which provides a packet-based non-line-of-sight broadband wireless access. The standard mainly extends the base 802.16-2004 to support mobility and energy-efficient sleep and idle modes suitable for mobile devices operated by energy-limited batteries. Both the 802.16-2004 and 802.16e-2005 standards support multiple advanced features such as orthogonal frequency division multiple access (OFDMA), multi-input multi-output MIMO with space-time coding and spatial multiplexing options, beam forming and adaptive antenna system (AAS), adaptive modulation and coding (AMC), time-division duplexing and frequency-division duplexing modes among others. This rich feature set is intended to provide data rates in excess of tens of megabits per second. As such, WiMAX should meet mobile Internet access requirements. It supports multiple handoff mechanisms, ranging from hard handoffs to soft handoffs, power-saving mechanisms for mobile devices, and advanced quality-of-service (QoS) and low latency for improved support of real-time applications with an optimized service class particularly for VoIP.

The IEEE 802.16e standard is complimented by interoperability and conformance specifications defined by the WiMAX Forum [4-6]. The commercial realizations of products based on the 802.16e standard are typically

referred to as mobile WiMAX (same as IEEE 802.11 products are commercially known by the name WiFi).

There are a few crucial factors that give WiMAX an edge over other competing wireless technologies, which include [7]:

1. Superior radio performance: this is largely tied to the use of OFDMA, a multiplexing technique well suited to multipath environments that gives network operators higher throughput and capacity, great flexibility in managing spectrum resources, and improved indoor coverage. In addition, support of various MIMO modes, beam forming, and hybrid ARQ also boost throughput the efficiency of the radio link.

2. Spectrum usage: Mobile WiMAX can be deployed in several licensed bands (2.3 GHz, 2.5 GHz, 3.3 GHz, 3.4-3.8 GHz) with channel sizes ranging from 3.5 MHz to 10 MHz. The availability of multiple FFT sizes fit different channel bandwidth. This gives operators the flexibility to use WiMAX in multiple spectral bands in the available spectrum they have.

3. Advanced IP-based architecture: Mobile WiMAX is based on IP-core and its protocol model is designed with a packet-transport model. Therefore, it should enable the transition to an all-IP network. In essence, the WiMAX Forum is working on supporting IP MultiMedia Subsystem (IMS) and its 3GPP2 counterpart, MultiMedia Domain (MMD).

4. Native protocol support of multiple services: Mobile WiMAX supports five classes of service that could serve multiple applications including circuit-like operations, VoIP, video streaming, and web traffic. The medium access control (MAC) layer also encapsulates functionality for traffic policing for enabling user-network traffic contracting and control. The following section provides details about the WiMAX QoS model.

In this paper, we focus on one of the important aspects of mobile WiMAX, namely the scheduling performance at the mobile stations. We consider the scheduling problem from two aspects. The first is satisfying the functional requirement in, sharing the allocated BW resource between different connections according to negotiated QoS parameters of each connection. The second is to introduce a computationally efficient algorithm to meet the real time constraints on typical mobile hardware platforms.

The rest of this paper is organized as follows. In section 2, we provide a brief overview of the quality of service model in WiMAX. Section 3 gives details about the proposed uplink scheduling. The corresponding results showing its excellent performance are presented in section 4. Details of the implementation on an ARM9 processor core, and real-time performance are provided in section 5.

Section 6 concludes the paper.

2. WiMAX Quality-of-Service Model

WiMAX layer 2 transport is a connection-oriented protocol. All data transports must be established in advance between the base station and mobile station. The service is based on a centralized media access and resource management scheme where all resources are centrally managed by the BS. The BS has complete information about the attached MS's and their established connections.

The uplink medium access is based on a request/grant mechanism, where the mobile station (MS) requests bandwidth from the base station (BS). The BS performs scheduling among the contending MSs connections and allocates grants to meet the connections' quality-of-service parameters. The downlink is a broadcast medium access. The BS has the full responsibility for forming the downlink subframe and filling it with the contending packets destined for the various MS's.

The BS transmits a broadcast message (ULMAP) in every frame to announce the uplink scheduled allocations. In order to minimize the overhead of the ULMAP message, the resources allocated for all connections of some MSs are aggregated and added as a single entry in the ULMAP message.

According to the IEEE standard the principal mechanism for providing QoS is to associate packets traversing the MAC interface (through the Convergence Sublayer (CS) SAP) into a service flow as identified by the Service Flow ID (SFID) and connection as identified by Connection ID (CID). A service flow is a unidirectional flow of packets that is provided a particular QoS. The MS and BS provide this QoS according to the QoS Parameter Set defined for the service. Therefore, in order to have a full picture of the WiMAX QoS model, we should carefully inspect two main concepts: service flows management and quality of service classes. This is explored in the following two sections.

2.1 Management of Service Flows and Connections

A Service Flow (SF) is a MAC transport service that provides unidirectional transport of packets on the uplink or on the downlink. A service flow is identified by a 32-bit service flow identifier (SFID). The service flow defines the QoS parameters for the packets (PDUs) that are exchanged on the associated connection. A connection is a MAC-level unidirectional connection between a BS and an MS or vice-versa. It is a unidirectional mapping between BS and MS MAC peers for the purpose of transporting a given service-flow's traffic. A connection is only for one type of service and is identified by a 16-bits connection identifier (CID).

A service flow can either be provisioned, admitted or active. A provisioned service flow is pre-defined by a network management system. A two-phase activation model is used in the standard to activate service flows (direct activation is also possible). In the two-phase activation

model, the resources for a flow are first 'admitted' and then, once the end-to-end negotiation is completed, the resources are 'activated'. A connection can only be associated with an admitted or active service flow. The mapping from connection to service flow is 1:1 whereas from service flow to connection it is either 1:0 (provisioned) or 1:1 (admitted/active).

The standard allows both the MS and BS to dynamically manage service flows. This is achieved via the management messages: DSA-REQ/RSP, DSC-REQ/RSP, and DSD-REQ/RSP (respectively dynamic service add request/response, dynamic service change request/response, and dynamic service delete request/response messages). Each of these messages service flow parameters and/or admission/de-admission or activation/deactivation is negotiated between the BS and the MS. The BS must approve all such transactions to provide better management of QoS and connection resources.

2.2 Quality of Service Classes and the Associated Parameters

WiMAX defines five quality of service classes:

- **Unsolicited grant service (UGS):** this service provides support for real-time data streams consisting of fixed-size data packets issued at periodic intervals such as uncompressed voice and TDM circuits. The base station is supposed to provide periodic grants for this type of service.
- **Real-time polling service (rtPS):** this service provides support for real-time data streams consisting of variable-sized data packets that are issued at periodic intervals, such as compressed video or voice telephony. The BS provides periodic polling grants for connections belonging to this type of service allowing them to request bandwidth.
- **Extended real-time polling service (ertPS):** this service provides support for real-time data streams consisting of variable-size data packets that are issued at periodic intervals, with activity detection, such as VoIP service with silence suppression. For such applications, using ertPS is a good compromise between low delay achieved by UGS and efficient BW utilization offered by rtPS [8-9].
- **Non-real-time polling service (nrtPS):** this service provides support for delay-tolerant data streams consisting of variable-size data packets for which a minimum data rate is required, for example web browsing.
- **Best-effort (BE):** this service supports data streams for which no minimum service level is required such as background file transfer or E-mail traffic.

As can be seen, these service classes are defined in a manner such as to support all kinds of scenarios/services

Table 1: Mandatory QoS parameters for the different service classes

Scheduling Type	UGS	ertPS	rtPS	nrtPS	BE
Maximum sustained traffic rate (MSTR)	Yes	Yes	Yes	Yes	Yes
Maximum burst size (MBS)	Recommended	Recommended	Recommended	Recommended	Recommended
Minimum reserved traffic rate (MRTR)	Yes ¹	Yes	Yes	Yes	
Maximum latency	Yes	Yes	Yes		
Tolerated jitter	Yes	Recommended			
Unsolicited grant Interval	Yes	Recommended			
Unsolicited polling interval			Recommended	Recommended	
Time base			Recommended	Recommended	
Traffic Priority				Yes	Yes
Request/Transmission Policy	Yes	Yes	Yes	Yes	Yes

that could be offered by service providers.

When a service flow is defined, it is associated with one of these service classes. Each service class has associated quality of service parameters. WiMAX defines many service parameters and each service class has some obligatory parameters as well as optional parameters that could be associated with a service flow. The main QoS parameters are (for a full list refer to section 11.13 of IEEE 802.16 standard [1-3]):

- Scheduling service type: this parameter is one of the classes (UGS, eRTPS, rtPS, nrtPS, BE).
- Maximum sustained traffic rate (MSTR): this is the peak rate of the service in bits per second.
- Maximum burst size (MBS): this is the maximum burst size that is allowed for this service in bytes. This parameter is used with the MSTR to apply the traffic policing.
- Minimum reserved traffic rate (MRTR): this is the minimum rate reserved for the service flow expressed in bits per second. The service is expected to always be able to transmit at this rate. The BS provides the proper admission control and resource management to guarantee this rate over time.
- Maximum latency: this is the maximum interval in milliseconds between the entry of a packet at CS of the BS or the MS and the reception of the packet at the peer CS. This service commitment does not have to be met for service flows that exceed their minimum reserved rate.
- Tolerated jitter: this is the maximum allowed delay variation in milliseconds.
- Unsolicited grant interval: this is the nominal interval between successive data grant opportunities in milliseconds.
- Unsolicited polling interval: this parameter specifies the maximal nominal interval between successive

polling grants opportunities for the service flow in milliseconds.

- Time base: a time window in milliseconds used for minimum rate measurements.
- Traffic priority: this parameter specifies the priority assigned to a service flow. This is only relevant for service flows that are of the same class and have identical QoS parameters. The higher priority service flows should be given preferential scheduling treatment.
- Request/transmission policy: the value of this parameter provides the capability to specify the allowed means for the associated service flow to request uplink bandwidth allocation from the BS. It also specifies whether the packing and fragmentation are allowed when forming PDU's of this service flow.

Table 1 specifies the binding between service classes and QoS parameters indicating which parameters are optional/mandatory for service classes.

3. The Proposed Mobile Station Uplink Scheduling Scheme

The proposed WiMAX Mobile Station Scheduler is a credit-based scheduling mechanism. The main complexity in the MS scheduling stems from the fact that while the MS sends a Bandwidth Request (BW) on a per-connection basis, the BS does not allocate the grant per MS on a connection basis. It is the responsibility of the MS to distribute this allocated grant among the connections. According to the scheduling algorithm, the share of the grant allocated to a particular connection could be different (either smaller or larger) from the BW requested earlier for that connection. When the received grant is inconsistent from the requests, then a simplified scheme that allocates bandwidth in say round-robin manner could cause unfairness and violation of the expected service quality of the connections. The proposed scheme is simple and scalable (can serve many connections), and has been shown

¹ For UGS connections maximum sustained traffic rate (MSTR) and minimum reserved traffic rate (MRTR) are equal.

to be able to satisfy real-time requirements. The scheme achieves the following:

- Isolation of all classes under normal operation and overload conditions.
- Prevention of high-priority traffic from stealing bandwidth of other traffic. All connections are provided with the amount of bandwidth they requested. Excess bandwidth is divided among connection according to their priority.
- Prevention of starvation for Best Effort traffic
- Support for five classes of connections (UGS, ertPS, rtPS, nrtPS, BE) in addition to management traffic.

In brief, the MS implements a scheduling module with the following inputs and outputs (Figure 1). This module shall be used in **every uplink sub-frame**.

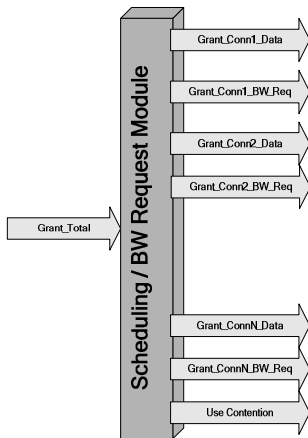


Figure 1: The scheduling and bandwidth request module

Inputs: Total Grant allocated for the MS in the uplink sub-frame. This grant may be zero or more bytes.

Outputs:

1. The part of grant to use for serving the data of each connection. That may be zero or more bytes.
2. The part of grant to use for sending a BW request for each connection. That may be zero or six bytes.
3. A decision to start using contention opportunities for sending BW request if needed.

The overall architecture is shown in Figure 2. The Uplink traffic is classified into proper connections and inserted into the connection queue. Each connection has its own dedicated queue and is subject to traffic policing. Management messages (basic, primary and secondary) each has its own connection queue. The focus of this paper is primarily on upper layer traffic connections, and shall not discuss the details of handling management messages.

In case of a frame containing UL allocations, an uplink MAP (UL-MAP) is received from the BS. The UL-MAP processing module parses the UL-MAP to obtain the

allocated data grant. It also constructs a transmission vector that describes the utilization of the different parts of the uplink suitable for transmission over the OFDMA-based physical layer. The scheduling module is then invoked to provide the portions of the grant allocated to the different connection and bandwidth requests. The PDU formation module then constructs the uplink PDU's from the connections queue, and could optionally apply fragmentation and packing if needed. This paper does not cover the details on the fragmentation, packing, and transmission vector formation. The focus of the following sections is limited to the scheduling aspects.

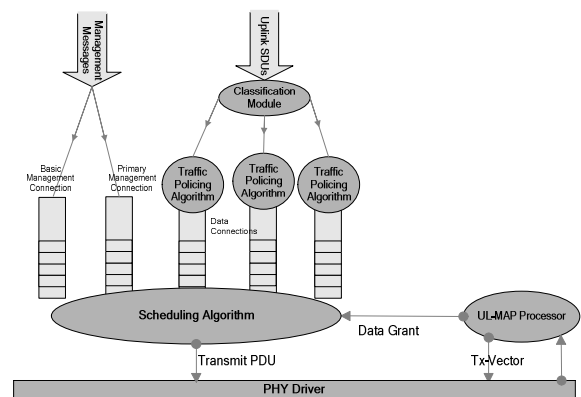


Figure 2: The scheduler architecture

3.1 Deficit-Round-Robin Scheduling

The proposed method is based on weighed deficit round robin (WDRR) scheduling. WDRR is a credit-based scheduling algorithm that is typically used to eliminate the processing overheads typically associated with methods such as weighted-fair queueing (WFQ) [10].

In a typical WDRR algorithm, the following procedure is followed [10]: each connection is assigned a quantum. This quantum is a function of the rate desired for the connection. When a grant is received, starting from the highest priority connection:

- The credit of the connection is incremented by one quantum.
- The connection is served as long as the size of the head-of-line (HoL) packet is smaller than the credit of the connection.
- The connection credit is decremented with the size of packets served.

WDRR is based on the concept that as long as the grant is not finished, the algorithm loops around all connections, starting from the first one, until the last connection is served. When the algorithm stops at a certain connection, the connection is granted bandwidth upto its current credit. This is continued until the grant is completely consumed. Typical WDRR assigns each connection a small quantum

and loops through connections multiple times. While that achieves a high level of fairness between different connections, looping for multiple times could degrade performance.

The operation of WDRR is shown in Figure 3. which depicts scheduling of four connections. The connections from top to down have quantum values of 30, 75, 75, 50 units. Initially, each connection has accumulated credit and packets with the shown sizes queued as depicted in the top of the figure. When the WDRR serves these connections, only the two middle connections have enough grants to serve the two HoL packets which results in their credit to decrease by the served packet size. In the next step, each connection accumulates more credit by adding one quantum to each connection.

The proposed scheduling method differs from the WDRR in the sense that it sets a limit on the number of rounds each connection can be served, as will be explained in section 3.2. In essence, only the first two rounds emulate WDRR. Moreover, the quantum used is not fixed but incremented with the a variable credit update, this variable credit update is a function of the minimum rate of the connection and the elapsed time since the connection credit was last updated.

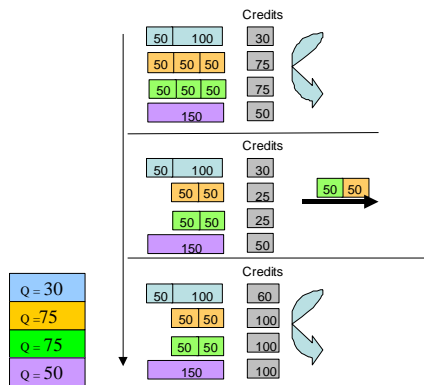


Figure 3: WDRR Operation

3.2 The Scheduling Rounds

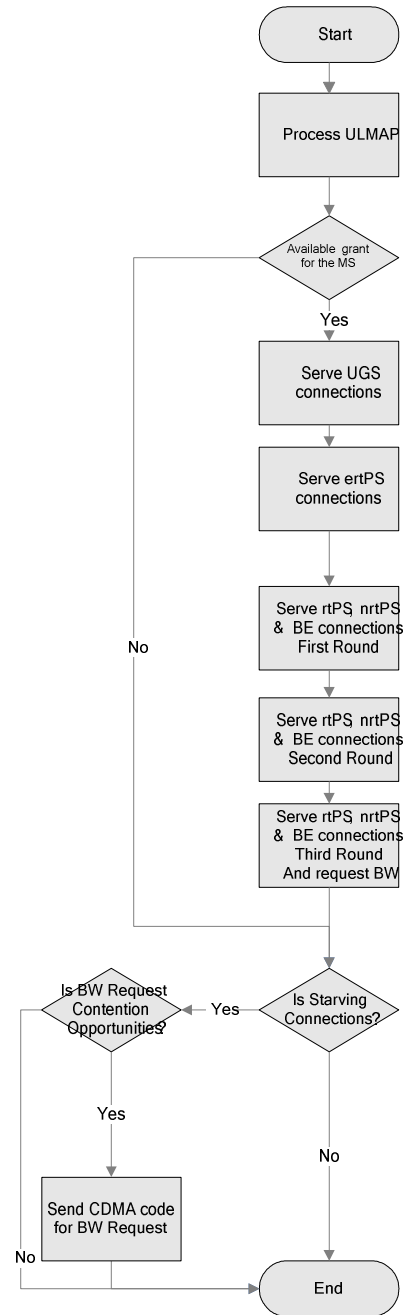


Figure 4: The overall scheduling scheme

The proposed scheduler starts with serving UGS connections in a single round. As the maximum sustained traffic rate is the same as the minimum reserved traffic rate for UGS connections, it is guaranteed that the BS shall allocate grant for all of the pending traffic of UGS connections. The scheduler serves ertPS connections in a similar manner.

Table 2: First Experiment - Parameters of the four connections

	Queue size (Pkts)	Pkct Rate (pkct/sec)	Pkct Length	Max latency (ms)	Min reserved rate (bps)	Max sustained rate (bps)	Frag.	Packing
UGS	200	292	uniform(93, 93)	20	148,800	148,800	FALSE	FALSE
rtPS	200	1171	uniform(55, 73)	20	512,000	600,000	FALSE	FALSE
nrtPS	200	500	uniform(80, 220)	300	512,000	600,000	TRUE	FALSE
BE	200	62 to 1062	uniform(100, 500)	100,000	1,200	1,660,000	TRUE	FALSE

Then the scheduler works in three rounds to serve the rtPS, nrtPS and BE connections. In each round, the scheduler scans the connections starting from the connection with highest serving priority to the lowest serving priority. In the first round, each connection is served up to the minimum of its requested bandwidth and current credit.

If there is remaining grant from the first round, the scheduler starts a second round. In the second round, each connection is served up to its requested bandwidth that has not been granted in the first round. The second round typically consumes the entire available grant. The restriction of limiting the allocated grant to a connection to the amount of requested bandwidth will prevent connections with high serving priority from stealing bandwidth requested previously by the lower priority connections.

If there are remaining grant after the second round, the scheduler proceeds to the third round, in which connections will be served up to whatever grant is available to serve the waiting traffic in the connection queues. In the third round, a pending SDU may be fragmented to fit with the remaining bandwidth if fragmentation is enabled on the connection. The overall scheme is depicted in Figure 4.

4. Performance Evaluation of the Proposed Uplink Scheduling Scheme

In this section we report results showing the excellent performance of the proposed uplink scheduling scheme. We use a discrete-event simulator which encapsulates the code for the MS upper MAC layer containing the scheduling code. Traffic is generated and time-stamped upon entry to the CS layer of the MAC. The system collects statistics on a per-connection basis such as average delay in the queues, average and distribution of the number of packets in the queue, number of dropped packets due to a full buffer at the connection queue, dropped packets due to policing, etc. The simulation is executed for a sufficiently large number of frames which in the shown results is taken as 100,000 frames. The uplink capacity is provisioned at a maximum

of 3 Mbps. All packets are generated in accordance with a Poisson process at a rate specified in the experiment parameters. This is true for all types except for UGS in which case, the packets are generated in accordance with a constant rate. The purpose of the simulation is to show the possibility of isolating traffic classes without impacting the low priority traffic.

We perform two experiments. In both, we use four connections of different types: UGS, rtPS, nrtPS, and BE. The MS is provisioned for a total uplink capacity of 3 Mbps. In the first experiment, the UGS, rtPS and nrtPS connection rates are fixed while the BE connection rate is varied. The purpose of the experiment is to show that with increasing BE traffic rates, other connections performance is hardly affected. In the second experiment, we maintain a fixed input bit rate for the UGS, nrtPS and BE connections while varying the rtPS connection rate. The purpose of this experiment would be to show that even when a high-priority traffic connection rate is increase, other connections performance is not affected even for BE traffic until exceeding overall provisioned capacity.

4.1 First Experiment

In the first experiment, the parameters of the four connections are as shown in Table 2.

We report the following performance metrics obtained as the BE connection rate is increased (the minimum reserved rate is fixed): average delay for each of the connections and the dropping rates (due to expiry, due to full buffers, and due to policing), in each connection. The overall BE rate is varied from 62 packets/sec to 1062 packets/sec resulting in an uplink subscription ration in the range of 50% to 130% when the uplink is set to a maximum rate of 3 Mbps.

Figure 5 shows the results of the average delay received by the three connections. We notice here that regardless of the overall load, the rtPS and nrtPS connections delays are not affected. The traffic delay starts to increase when the subscription ratio surpasses 95%. Figure 6 also shows that the dropping percentage in the BE SDU traffic due to a full

Table 3: Second Experiment - Parameters of the four connections

	Queue size (Pkts)	Pckt Rate (pkt/sec)	Packet Length (bytes)	Max latency (ms)	Min reserved rate (bps)	Max sustained rate (bps)	Frag.	Packing
UGS	200	200	uniform(93, 93)	20	148,800	148,800	FALSE	FALSE
rtPS	200	293 to 5566	uniform(55, 73)	20	150,000 to 1,950,000	2,250,000	FALSE	FALSE
nrtPS	200	500	uniform(80, 220)	300	512,000	600,000	TRUE	FALSE
BE	200	125	uniform(100, 500)	100,000	1,200	300,000	TRUE	FALSE

queue starts to increase as the BE load increases. No noticeable drops in the UGS, rtPS or nrtPS connection queues are observed.

4.2 Second Experiment

The parameters of the four connections are shown in Table 3.

We report the same performance metrics as experiment one as both the rtPS connection rate and its minimum reserved traffic rate are increased. The connection rates are selected to vary the link subscription ratio from 40% to 130%. We expect that all connections have their minimum rate met until the link utilization approaches 100% (i.e. overload condition).

Figure 7 shows the average delay for each of the connections. The rtPS connection performance remains consistent for all utilization levels. However, by inspecting Figure 8, we can see that when the incoming rate increases above 3 Mbps, rtPS traffic starts to be dropped due to expirations. The nrtPS and BE connections have good performance. The delays are constant until the overall incoming rate increases above 3.15 Mbps where the delays increase. In Figure 8, we notice BE packet dropping due to full queue for all incoming rates between 0.2 and 1.9%. The dropping in rtPS (when the incoming rate increases above 3 Mbps) keeps the performance predictable for all classes. When rtPS rate increases above its maximum sustained traffic rate, policing starts to limit the rtPS rate and the delay performance remain predictable.

5. Implementation of the Scheduler on ARM9 Platform

The target for possible incarnation of the proposed scheme is in a WiMAX mobile stations MAC layer firmware. Typically such firmware is implemented on SoC via ASIC or embedded processors available as core IP such

as ARM [11] and PowerPC [12]. In this section we report real-time results of implementing the proposed scheduler on an ARM9 platform.

We select the TI OMAP 5912 [13] development kit which is a popular platform for developing mobile handsets since it has both a TI TMS320C55x DSP and an ARM9 general purpose processor. The ARM9 has a data cache of 8 Kbytes, an instruction cache of 16 Kbytes, and runs on a clock of 192 MHz. The board comes with a real-time operating system (RTOS) that is sufficient to evaluate and assess the real-time performance of the scheduler and show that the design is implementable and suitable for the target embedded devices market.

In the experiments, simultaneous uplink and downlink traffic are generated. The uplink traffic is provisioned at a maximum rate of 3.2 Mbps. The downlink traffic is about 10 Mbps. In each experiment, we measure the rate allocated to each of the connection after 10, 20, and 30 seconds as a means to measure the stability of the scheduling scheme performance. Three different experiments were conducted to measure the CPU utilization of the ARM9 and stability of the scheme.

In the first experiment, we had one UGS connection and a BE connection with a total incoming uplink rate of 3.2 Mbps. The CPU utilization is observed at 78%.

In the second experiment, we maintained the maximum rate of 3.2 Mbps while five connections with different scheduling types were running. The CPU utilization slightly increased to 79.4%.

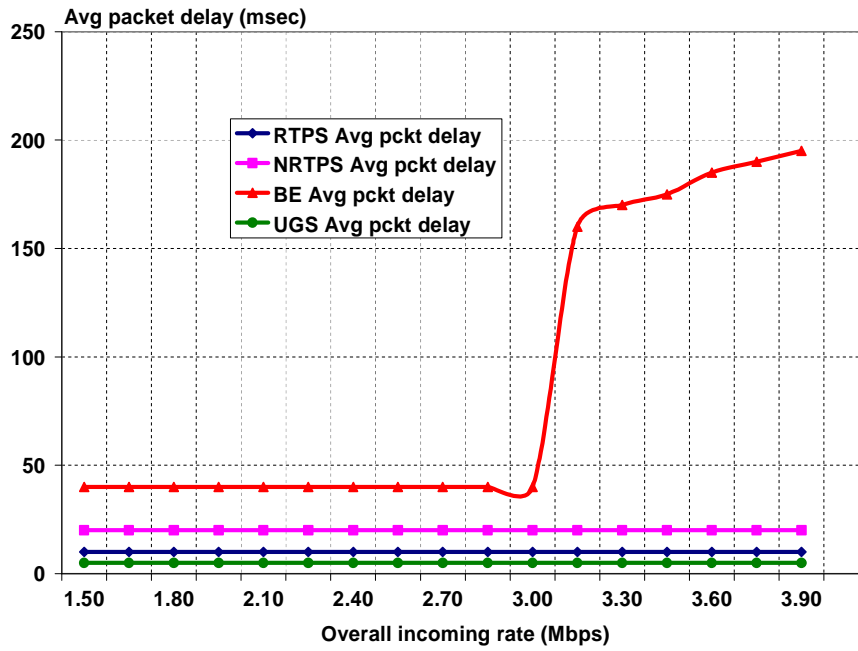


Figure 5: First Experiment - Delay Performance for the Connections

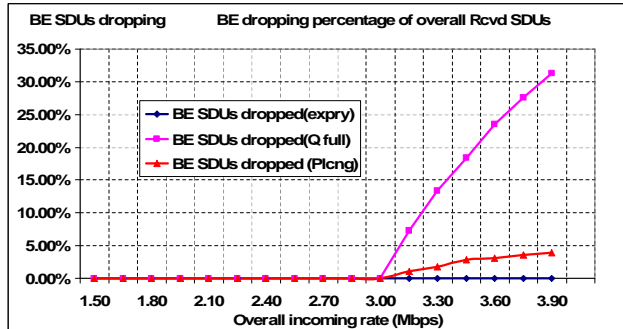
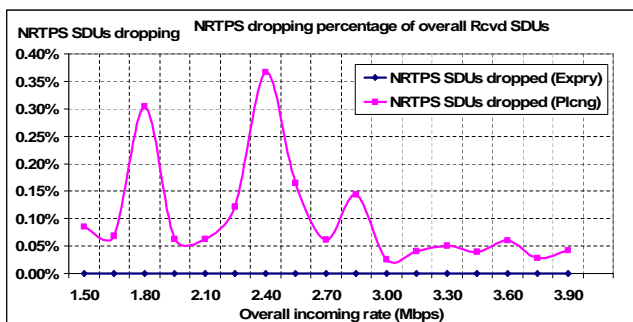
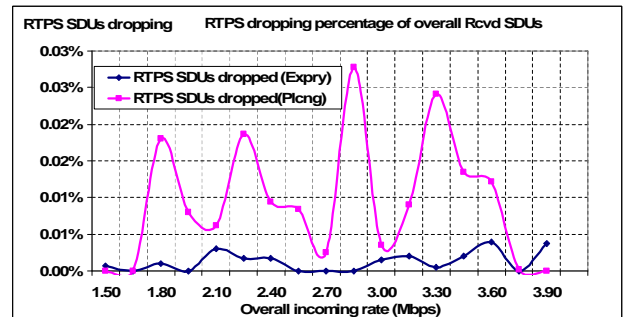
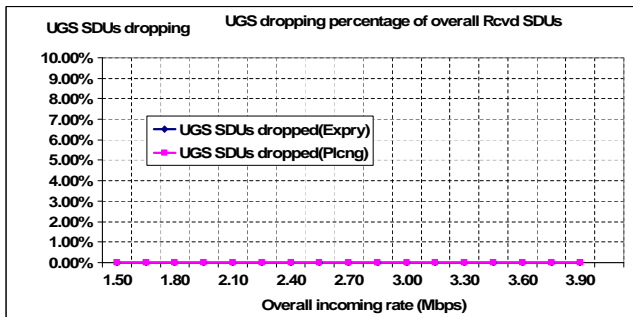


Figure 6: First Experiment – SDU dropping rate for the UGS, rtPS, nrtPS, and BE connections

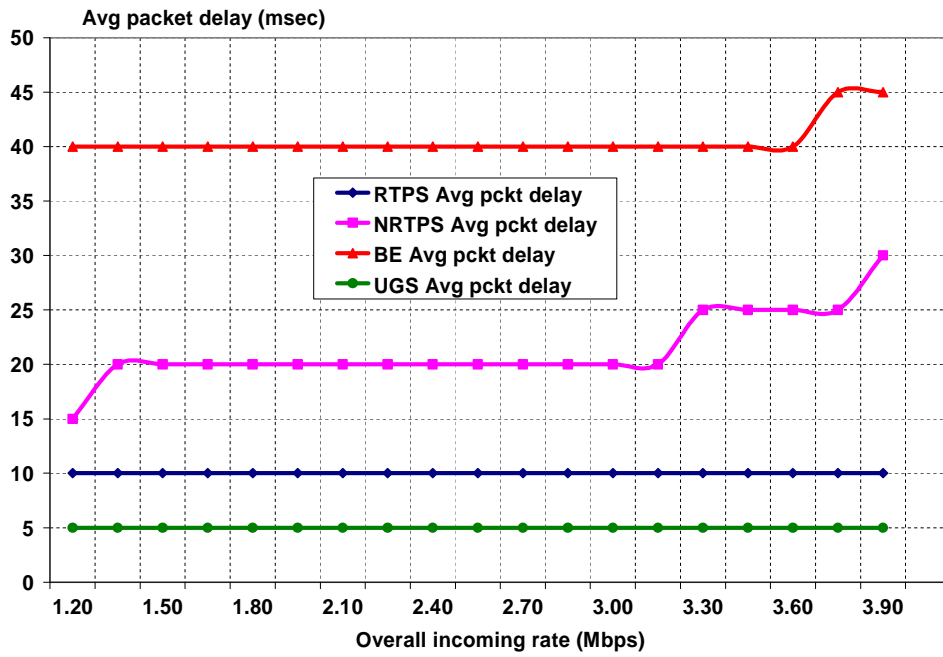


Figure 7: Second Experiment - Delay Performance for the Connections

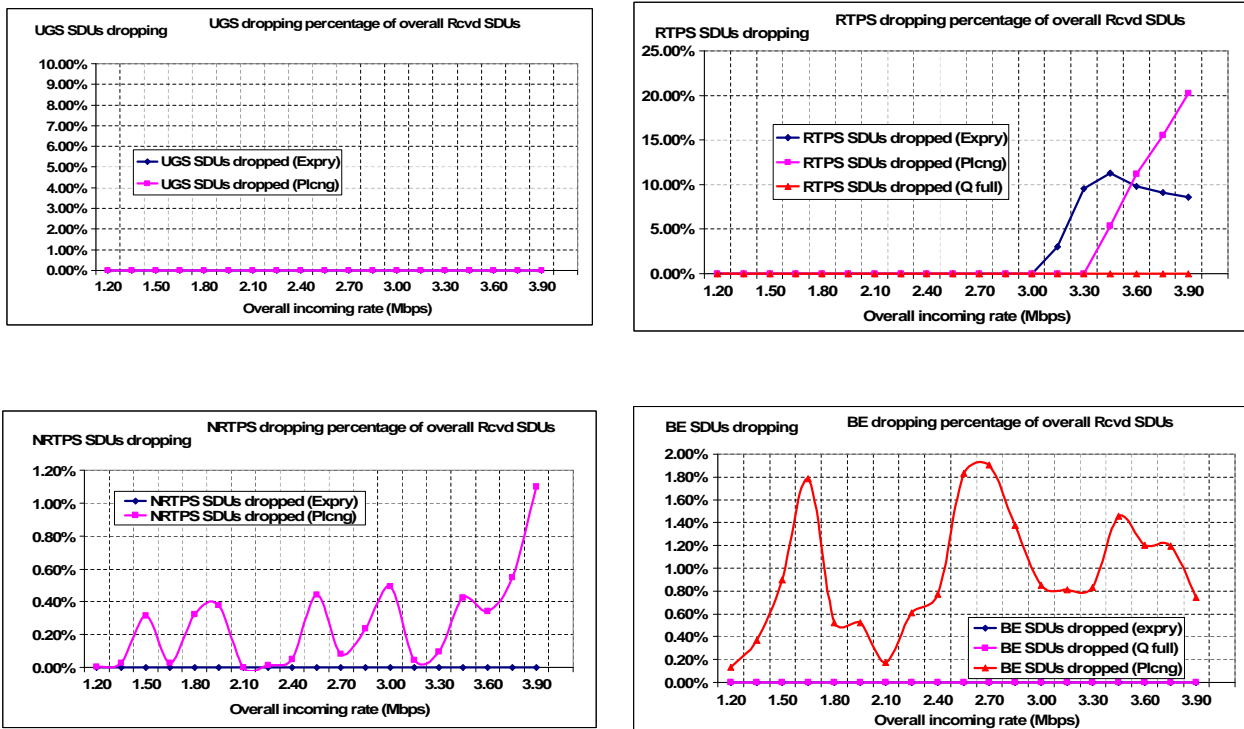


Figure 8: Second Experiment – SDU dropping rate for the UGS, rtPS, nrtPS, and BE connections

Table 4: Real-time performance of the scheduler under different loading on ARM9

Type	%	Assigned BW (Mbps)	Packet/10ms	Test Rates (Mbps) (percent of Total BW)			CPU Utilization
				After 10 sec	After 20 sec	After 30 sec	
Experiment (1)							
UGS	40%	1.28	4	1.28 (40%)	1.2799 (39.99%)	1.2795 (39.98%)	78.00%
BE	60%	1.92	6	1.92 (60%)	1.9190 (59.97%)	1.9187 (59.95%)	
Experiment (2)							
UGS	20%	0.64	2	0.64 (20%)	0.6394 (19.98%)	0.6391 (19.97%)	79.40%
BE	40%	1.28	4	1.28 (40%)	1.2787 (39.95%)	1.2783 (39.94%)	
RTPS	10%	0.32	1	0.32 (10%)	0.3197 (9.99%)	0.3196 (9.98%)	
NRTPS	10%	0.32	1	0.32 (10%)	0.3197 (9.99%)	0.3196 (9.98%)	
ERTPS	10%	0.32	1	0.32 (10%)	0.3197 (9.99%)	0.3196 (9.98%)	
Experiment (3)							
UGS	30%	0.96	3	0.9610 (30.03%)	0.96 (30%)	0.96 (30%)	82.30%
BE	70%	2.24	7	0.3206 (10.02%)	0.32096 (10.03%)	0.3212 (10.03%)	
RTPS	20%	0.64	2	0.64 (20%)	0.6397 (19.99%)	0.6398 (19.99%)	
NRTPS	20%	0.64	2	0.64 (20%)	0.6397 (19.99%)	0.6398 (19.99%)	
ERTPS	20%	0.64	2	0.64 (20%)	0.6397 (19.99%)	0.6398 (19.99%)	

In the third experiment, we had an overload scenario where the offered load exceeds 5 Mbps with the total maximum uplink rate still at 3.2 Mbps. The CPU utilization increased to 82%, which was still adequate to meet the required real time constraints. The ARM9 was capable of meeting the 5 msec frame duration constraint, and all scheduling for a specific frame was concluded before the arrival of the following frame.

6. Conclusions

The WiMAX mobile station uplink scheduling has been studied and optimized from two aspects. The first is to meet the functional requirements, and the second is to meet the real-time constraints. The functional requirement is defined as sharing the allocated grant from the BS between different connections in a fair manner according to each connection QoS parameters. The real-time constraint is to perform scheduling and prepare the PDU to transmit in the limited interval between receiving the UL-MAP and the transmission time. Experimental results demonstrate that the proposed algorithm meets both the functional requirements even in extreme overload conditions and the real-time constraints.

References

[1] IEEE WirelessMAN 802.16 Group, Draft IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems. IEEE P802.16-REVd/D5-2004.
 [2] IEEE WirelessMAN 802.16 Group, Draft IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. Amendment for Physical and Medium Access

Control Layers for Combined Fixed and Mobile Operation in Licensed Bands. IEEE P802.16e/D11, September 2005.
 [3] IEEE WirelessMAN 802.16 Group, Draft IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. Corrigendum 2. IEEE P802.16-2004/Cor2/D3, March 2007.
 [4] WiMAX Forum, “Mobile Test Suite Structure and Test Purposes (TSS&TP) Specification Revision 2.0.0,” Aug. 2007.
 [5] WiMAX Forum, “Mobile InterOperability Test (MIOT) Revision 2.0.1,” Aug. 2007.
 [6] WiMAX Forum, “Mobile Radio Conformance Tests (MRCT) Revision 1.1.0,” July 2007.
 [7] Wimax Forum, “Mobile WiMAX: The Best Personal Broadband Experience!,” June 2006.
 [8] Howon Lee, T. Kwon, D. Cho, “An Enhanced Uplink Scheduling Algorithm Based on Voice Activity for VoIP Services in IEEE 802.16d/e System,” IEEE Communications Letters, pp. 216-218, Aug. 2005.
 [9] Howon Lee, Taesoo Kwon and Dong-Ho Cho, “ Extended-rtPS Algorithm for VoIP Services in IEEE 802.16 systems,” Proc. of the IEEE ICC 2006, pp. 2060-2066, June 2006.
 [10] S. Keshav, “ An Engineering Approach to Computer Networking”, Addison Wesley Longman, Inc., 1997.
 [11] ARM INC, ARM Processor Families, <http://www.arm.com/products/CPUs/families.html>.
 [12] IBM, PowerPC Cores, http://www.ibm.com/chips/techlib/techlib.nsf/products/Power_PC_Cores.
 [13] Texas Instruments, OMAP Platform, <http://www.ti.com/omap>.