A Switched Scheduling Algorithm for Congestion Relief in WiMAX Wireless Networks

Ahmed H.El-Shinnawy, Amin M.Nassar
Faculty of Engineering
Cairo University
ahmedhee@aucegypt.edu, aminassar@gawab.com

Ashraf H.Badawi
Wimax Program Office
Intel Corportation
ashraf.h.badawi@intel.com

***Abstract*—This paper focuses on improving the scheduling capabilities of the IEEE 802.16-2005 broadband wireless networks, through utilizing multi-queuing algorithms and switching between those queuing algorithms over time to maximize the performance of the WiMAX system. The proposed algorithm aims to improve throughput of high priority traffic in congested networks, without violating the right of lower-priority-traffic to be served in adequate manner.**

In this work, a detailed simulation study is carried out for the proposed scheduling algorithm as well as comparing its performance with some known algorithms such as Proportional Fairness (PF)[12], Round Robin (RR), and Strict-Priority. Analyses and evaluation of the performance of the scheduler to support the different QoS classes is given as well. The simulation is carried out via the Opnet modeler simulator [13]. The results show the proposed algorithm is capable to handle different users' requirements under congestion conditions.

Keywords-QoS; Opnet; OFDMA; Scheduling Algorithms; WiMAX

I. INTRODUCTION

Broadband wireless access (BWA) systems, [1], [2] are very flexible and easily deployable high-speed communication systems. BWA systems complement existing last mile wired networks such as cable modem and xDSL. IEEE 802.16 group aims to unify BWA solutions [1]. A technical overview of IEEE 802.16 is provided in[1], [3]. The objective is to have an efficient use of radio resources while serving different types of data flows. These flows can have different constraints such as minimum traffic rate, maximum latency, and tolerated jitter.

The IEEE 802.16-2005 standard supports three different physical layers,as shown in Figure(1): 1) Single Carrier, 2) OFDM/TDMA and 3) OFDMA [1]. OFDMA physical layer is the most efficient and complex one[4]. In OFDMA each substation (SS) can receive some portions of the allocation for the combination of time and frequency so that the channel capacity is efficiently utilized. OFDMA outperforms the OFDM & SC[4]. This research focuses only on OFDMA.

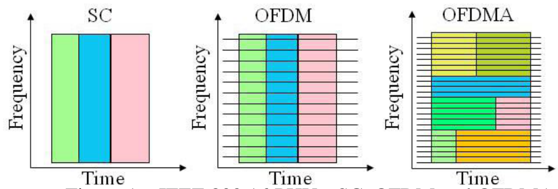

Figure 1. IEEE 802.16 PHYs: SC, OFDM and OFDMA

To support the different types of traffic with their various requirements IEEE 802.16-2005 defines five QoS service classes: Unsolicited Grant Scheme (UGS), Extended Real Time Polling Service (ertPS), Real Time Polling Service (rtPS), Non Real Time Polling Service (nrtPS) and Best Effort Service (BE).

UGS is designed to support real time data stream consisting of fixed size data packets issued at periodic intervals such as E1/T1 and voice over IP without silence suppression. The main QoS parameters are maximum sustained rate (MST), maximum latency and tolerated jitter (the maximum delay variation.

rtPS: This service class is for variable bit rate (VBR) realtime traffic such as MPEG compressed video. Unlike UGS, rtPS bandwidth requirements vary and so in this service, the BS provides periodic unicast (uplink) request opportunities, which meet the flow's real-time needs and allow the SS to specify the size of the desired grant. The QoS parameters are similar to the UGS but minimum reserved traffic rate and maximum sustained traffic rate need to be specified separately[4].

ertPS: This service class is designed to support VBR applications that have data rate and delay requirements, like the case in VOIP without silence suppression. The QoS parameters are the same as those in UGS[5].

nrtPS: This service class is for non-real-time VBR traffic with no delay guarantee. Only minimum rate is guaranteed. In the nrtPS scheduling service, the BS provide unicast uplink request polls on a 'regular' basis, one second or less, which guarantees that the service flow receives request opportunities even during network congestion.. In addition, the SS is allowed to use contention request opportunities. File Transfer Protocol

(FTP) traffic is an example of applications using this service class[4].

BE: This class is designed to support data streams for which no minimum service guarantees are required, like the case in HTTP traffic. The BS does not have any unicast uplink request polling obligation for BE SSs. Therefore, a long period can run without transmitting any BE packets [5]

In IEEE 802.16-2005, the BS (Base Station) centrally allocates the channels in different slots to different SSs (Subscriber Stations) for uplink and downlink. SSs in turn allocate these resources to the various connections they are supporting at that time. The process in the MAC access layers which is responsible for resource allocation is called the scheduling process. Unlike other parts of IEEE 802.16, scheduling was left for research to specify it. The optimal scheduling algorithm is still in open research area [6], [7], [8]. In this research a new scheduling algorithm is proposed to allow for better allocation of resources to different SSs in case of congestion, based on their QoS parameters and priority.

The reminder of this paper is organized as follows. Section 2 provides a review for relevant work and problem formulation. In section 3 a detailed description of the proposed algorithm is given. Section 4 describes setup of the simulation environment Section 5 shows the results and output of simulation of the algorithm compared to other scheduling algorithms. Concluding remarks and directions for future work are given in section 6

II. RELATED WORK AND PROBLEM FORMULATION

A. *Related work*

Recently published scheduling techniques for WiMAX can be classified into two main categories: channel- unaware schedulers and channel-aware schedulers[4]. Channel-unaware schedulers use no information of the channel state condition in making the scheduling decision. The design of those schedulers varies based on the ultimate goal of the scheduler like, maximizing throughput or fair allocation of resources between different SSs. However, the main challenge facing researches is the distinctive characters of each of the QoS classes. No single queue algorithm can handle all different types of parameters constrains simultaneously. For instance, no published researches show how to handle jitter over WiMAX, and most researches focuses on throughput rate or delay[4].

To avoid the obstacle above, many researches use Intra-class scheduling, where each class has a distinctive resource allocation mechanism that matches the requirements of the quality of service. Relation between inter classes is organized either on the bases of class-priority, where classes are served in the order UGS, ertPS, rtPS, nrtPS, BE, or on the bases of flow-priority, where highest priority flow is served first regardless of its class.

Even within the same class, there are many constrains that can't be handled through one scheduler[4]. For instance, rtPS class is defined, by maximum sustained rate, minimum reserved rate, maximum latency and priority. Most existing scheduling algorithms give precedence to flows based on one or two of those parameters and neglect the remaining. For instance, RR, and all its variations WRR, WDRR[11], focuses only on the distribution of resource over all queues regardless of flow priority nor its minimum rate requirements. Other technique, like Max-Min Fair Allocation, focuses on giving priority to lowest data rate flow regardless of its priority or delay constrains. On the other hand, strict priority algorithms put more emphasis on high priority traffic, but causes other lower priority flows to starve.

To sum up, since the primary goal of a WiMAX scheduler is to assure the QoS requirements, the scheduler needs to support at least the five basic classes of services with QoS assurance. Since it is very difficult for any scheduler to handle all the parameters in one step, a new technique is proposed here to switch between different scheduling criteria's so the scheduler can consider many parameters simultaneously. For the time being, the design of the new algorithm is considering handling three types of quening,:1) priority queuing, 2) Proportional Fairness, and 3) Minimum traffic Maximization

B. *Problem formulation*

This research aims to design a scheduling algorithm which is capable of maximizing throughput of priority traffic in congested domains, where bandwidth needs of SSs exceeds system capacity. Meanwhile, the algorithm considers serving less priority traffic in a fair manner. Fairness here is defined for less priority traffic as its ability to fulfill the class constrains, including Minimum Reserved Rate (MRR) and maximum latency without bandwidth reservation.

The research focuses on handling rtPS & nrtPS classes, UGS and ertPS classes are very sensitive for delay/bandwidth variation and will be handled via strict priority, where their needs are served before any other class. On the other hand BE class of service is not considered, since the standard doesn't specify any minimum requirements for its QoS parameters therefore BE classes will use the residual bandwidth.

III. PROPOSED ALGORITHM

As indicated above, most scheduling algorithms focus on one aspect of scheduling only (For example.

Fairness, starving avoidance, Priority, etc..). Therefore, this research focuses on the implementation of a new scheduling algorithm that is capable of considering multi QoS aspects and has the pros of the above scheduling algorithms, while avoiding their cons.

To handle many constrains, the proposed algorithm needs to switch between multi-queuing systems rather than adherence to a single one. Achieving this successfully requires a definition of the purpose of each queuing algorithm used and conditions to switch between the algorithms.

To achieve the above-mentioned criteria, the new algorithm defines two layers of priority, First layer, with higher precedence and always considered first, priority is given for any flow that need Bandwidth to achieve the minimum requirements of its class of service.(For example to achieve minimum bandwidth requirements of the class or not to violate maximum latency of a packet).The second layer of priority is given to priority traffic that needs to maximize its throughput.

.

Serving flows via the proposed algorithm is done as follows, and as illustrated in figure (2),

1- Ignore empty flows
2- Measure the time of packets waiting for service in each queue.
3- If one of queues' waiting time for service is more than 80% of maximum delay, serve queue at once
4- If all queues delay is less than 80% of a maximum delay, check queue for transmission rate.
5- Serve queue that didn't reach its minimums sustained rate, given priority to queue with lower transmission rate compared to minimum rate.
6- If all queues satisfied their MST, and delay constraints, priority is give to queues with high priority till they satisfy their Maximum sustained rate.

In the proposed flow of the algorithm, delay checking step is considered the most significant step for the success, of the algorithm. Without this step the time between waiting time between two packets in same flow might vary too much and violates flows constrains especially for RTP flows.

The algorithm also uses another technique to improve jitter and delay, which is dividing MST equally into smaller windows, typically 20 window, and each flow will not be allowed to send traffic bursts more than size of this window, which is equal MST divided by number of windows per second.

So the above technique will allow the algorithm to give priority to flows that need to be served to satisfy their minimum service requirements, while higher priority flows will the chance to reach their MST in case of lower priority flows have fulfilled their minimum needs.

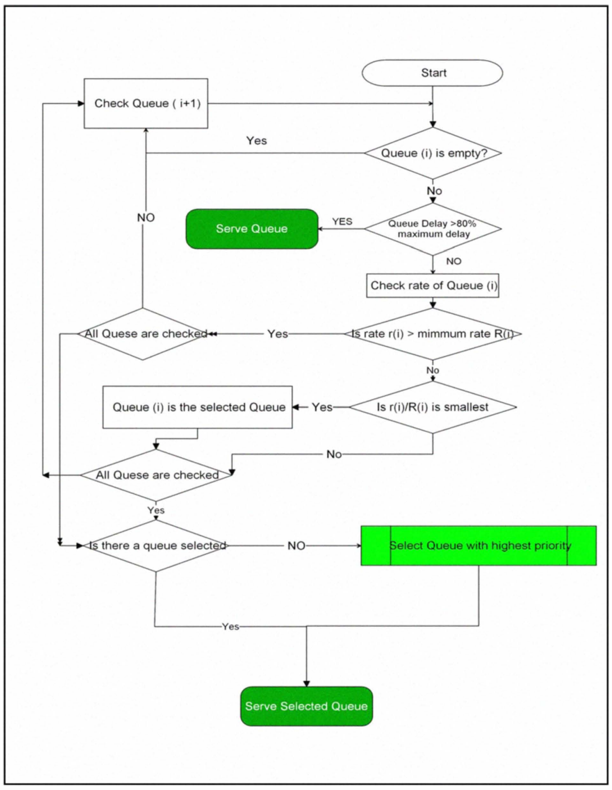

Figure 2. Flow Chart of the algorithm

IV. SIMULATION MODEL

The over all goal of the simulation model is to analyze the behavior and performance of the proposed algorithm in a congested domain The simulations have been performed using Opnet Modeler version 15.0 [13].The important parameters used to configure the PHY and MAC layers are summarized in table (1)

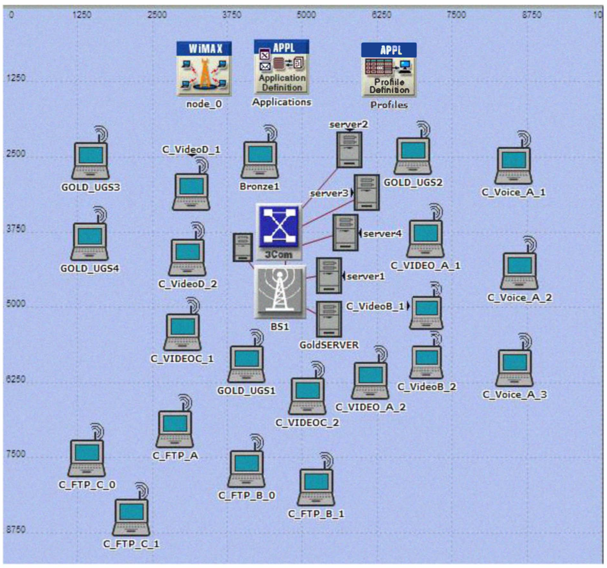

Figure 3. System Model Implementation in Opnet

The simulation assumes error-free channel since it makes it easier to prove assurance of QoS. Maximum theoretical capacity of the upload system is estimated as follows:

Upload Data rate = *number of uncoded bits per data symbol * total number of upload symbols*

.In this model : *number of uncoded bits per data symbol,= 560 * 6 *3/4 = 2520 bit*, where 560 is total number of data sub carrier for upload PUSC usage mode.

Data rate *=2520*12 symbol per frame *200 Frame= 6.048Mbps*

TABLE I. MAIN PARAMETERS OF THE SIMULATION MODEL

Model	Point to Multipoint
WIMAX channel bandwidth	= 10 MHz
Frame duration	5ms
Symbol Duration	102.86 Micro second
N	28/25
Delta_f	10.94khz
Number of sub carriers	1024
Frame structure	
Preamble symbols	1 symbol
Dublexing technique	TDD
Base Frequency	2.5GHZ
TTG	106 micro second
RTG	60 micros second
UL/DL Boundary	Fixed
UL sub frame size	= 12 slot
DL sub frame size	= 32 slot
Initial ranging	= 2 slot * 6 sub channel
Contention slot	= 1slot * 6 sub channel
Initial coding rate	¾
Initial modulation	64QAM

TABLE II. SERVICE FLOWS

Class	Max. rate Kbps	Min. Rate kbps	Pri	Traffic Type	No of SS
UGS	100	100	N/A	Video	2
RTP	384 ~600	200	20	Video	2~8
RTP	384	200s	10	Video	2
nRTP	384	200	20	FTP	1
nRTP	384	200	10	FTP	2
nRTP	200	100s	10	FTP	2
RTP	200	100	20	Video Conf.	2
RTP	200	100	10	Video Conf.	2
RTP	60	40	20	VOIP	3
BE	384	N/A	N/A	HTTP	2

The simulation environment consists of one BS and (20~26) SSs operating in IEEE 802.16 PMP mode. There will be one service flow between each SS and the BS. Traffic flows classes and their configuration is indicated in Table(2) and shown in Figure(3).

Since the simulated rtPS and nrtPS SSs are using polling service, which uses BPSK modulation at1/2 coding rate, it can be assumed that the idle average throughput of upload bandwidth is 5.5Mbps. *So,* the congestion criteria in this model is achieved via increasing the total maximum sustained rate requirements of all substations to exceed 5.5Mbps.

The congestion condition in the system is conducted via increasing throughput load on the BS by increasing number of flows of a specific type (for example video priority traffic).

V. RESULTS

In this section, the output of simulation is shown and analyzed .

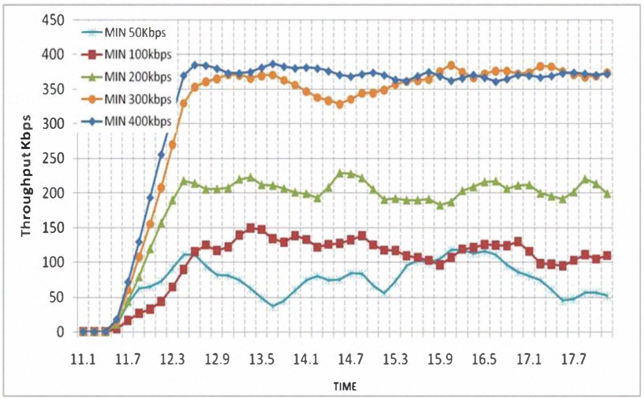

Figure 4. Average throughput of low priority traffic according to it MRR (Congestion 120%).

Figure (4) shows the throughput of number of low priority SSs. Each of the SSs has a different minimum reserved rate. The figures shows the performance of these stations at congestion level 120% as shown in the figure the proposed algorithm is successfully capable to meet the minimum reserved rate(MRR) for all SSs. Figure (5) shows that the algorithm is maintaining this performance even at very high congestion levels.up to 140%, as shown in the figure. It is shown in the figure that SS can have a throughput of 50% of its MST, which is equal to its MRR for this specific SS. The degraded throughput of Session in Figure 5 compared to RR and FQ is due to the fact that algorithm focuses on meeting the MRR for low-priority traffic, and preserves the remaining system bandwidth to boost the performance of higher-priority traffic as shown in Figure (6) & (7).

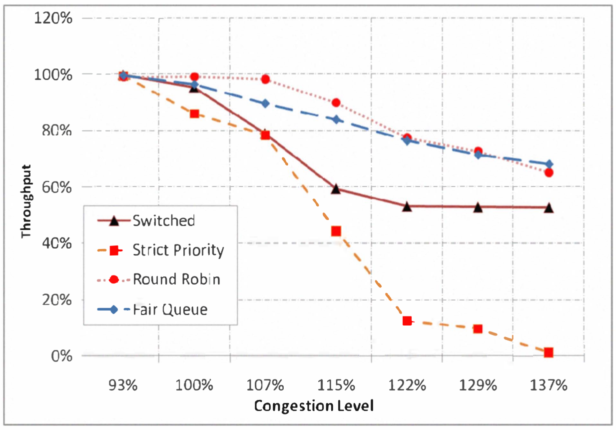

Figure 5. Throughput of low-priority traffic under different scheduling algorithms

On the other hand, it can be shown from Figure (6), that the proposed protocol doesn't just preserve the lower class constrains, MRR in this case, but it also posts the performance of the higher priority traffic. In the Figure it is shown that that all high priority traffic has much higher throughput compared to low-priority traffic regardless of the class lower limit. The figure shows that this is not the only advantage the algorithm provides. Another advantage is the flexibility the algorithm provides for service provider to boost the performance of high-priority traffic through controlling the MRR. The figure shows that high priority traffic is always aiming to achieve its MST, however if we can discriminate between two high-priority sessions through providing higher MRR for one of them. As shown in the figure, if one of the session has MRR very near or equal to the MST, the session will have much better chance to reach 100% of its MST compared to other high priority SSs.

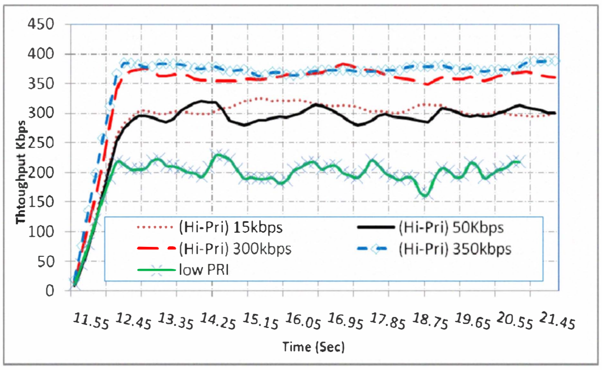

Figure 6. throughput of high priority traffic with different MMR compared to low-priority traffic.

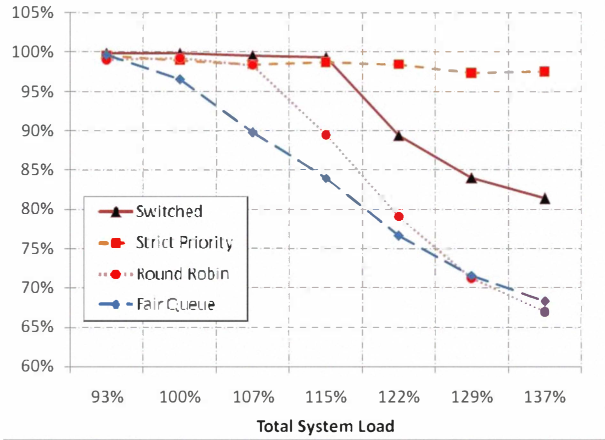

Figure 7. Throughput of High-priority traffic under different scheduling algorithms

Figure (7) compares the performance of the proposed switched algorithm against some well known algorithms under different congestion levels. The figure shows how the proposed algorithm has higher performance compared to other algorithm except of strict priority algorithm which has a superior performance for priority traffic but on account of low priority traffic which may starve under strict priority algorithm as shown in Figure (5).

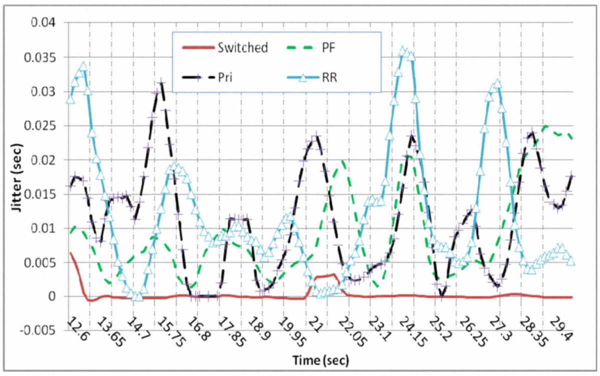

Figure 8. Jitter of VOIP sessions under different algorithms.

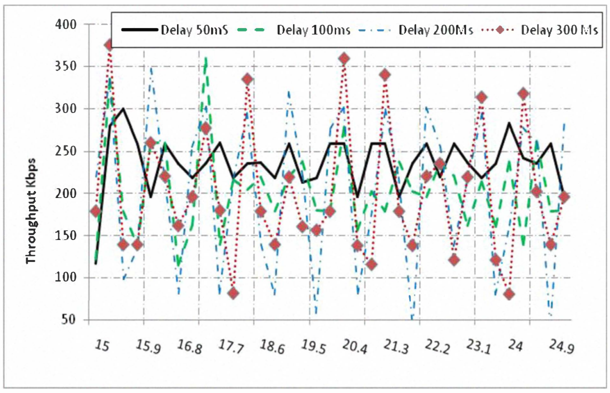

Figure 9. Controlling session throughput through changing maximum delay.

Figure(8) shows the effect of different algorithms on the jitter of VOIP session. It is shown that the proposed algorithm has much higher performance compared to other algorithms. The proposed switched algorithm has a higher performance since it is the only algorithm, through the shown ones, which considers delay as a parameter for taking its scheduling decision. The maximum delay constrains in the algorithm design makes the algorithms injects delay sensitive data, with low maximum latency < 50ms, at a semi-constant rate, which minimize the jitter or delay variation to minimum level. This options can be shown clearly in Figure 9, where the effect of changing the maximum latency of low priority traffic on session throughput is shown, It is clearly shown, that despite, the session in different latency cases has almost the same average throughput .the session with 50ms delay is injected at constant rate, with a very small variation between maximum and minimum throughput.

VI. CONCLUSIONS

In this paper, a new technique for scheduling in IEEE 802.16 Wireless MAN in PMP mode is proposed. The technique shows how scheduling flexibility and capabilities can be improved through switching between multi-queuing algorithms instead of fastening one queuing method only. The shown algorithm helps the BS to balance between serving high and low priority traffic simultaneously. The algorithm gives precedence to high priority traffic, to reach its MST, but only when all traffic sessions are capable to realize their lower class constrains like MRR and maximum latency. The algorithm also monitors all traffic flows to guarantee that no flow, whether has high or low priority will violate maximum delay limitation of the flow which make it excellent for delay sensitive applications like VOIP and video streaming.

REFERENCES

[1] IEEE P802.16Rev2/D2, "DRAFT Standard for Local and metropolitan area networks," Part 16: Air Interface for Broadband Wireless Access Systems, Dec. 2007

[2] IEEE 802.16 Working Group on Broadband Wireless Access. http://wirelessman.org

[3] C. Eklude, R. B. Marks, K. L. Stanwood and S. Wang. "IEEE Standard 802.16: A Technical Overview of the wireless MANTM Air Interface for Broadband Wireless Access," IEEE Communication Magazine June 2002

[4] Chakchai So-In, Raj Jain, Abdel-Karim Tamimi, "Scheduling in IEEE 802.16e Mobile WiMAX Networks: Key Issues and a Survey," IEEE Journal on Special Areas in Communications (JSAC), Vol. 27, No. 2, February 2009

[5] Loutfi Nuaymi, WiMAX TECHNOLOGY FOR BROADBAND WIRELESS ACCESS, John Wiley & Sons,2007

[6] L. Lin, W. Jia, and W. Lu, "Performance Analysis of IEEE 802.16 Multicast and Broadcast Polling based Bandwidth Request," in Proc. IEEE Wireless Communication and Networking Conf., Hong Kong, 2007, pp. 1854-1859.

[7] B. Chang and C. Chou, "Analytical Modeling of Contention-Based Bandwidth Request Mechanism in IEEE 802.16 Wireless Network," IEEE Trans. Veh. Technol., vol. 57, pp 3094-3107, Sept. 2008

[8] V. Alexey, Z. Ying, N. Qiang, and L. Andrey, "Efficient Request Mechanism Usage in IEEE 802.16," in Proc. IEEE Global Telecommunications Conf., San Francisco, CA, 2006, pp. 1-5

[9] Vandana Singh, Vinod Sharma, "Efficient and fair scheduling of uplink and downlink in IEEE 802.16 OFDMA networks", Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE,

[10] Ahmed H. Rashwan, Hesham M. El-Badawy, Hazem H. Ali, "Comparative Assessments for Different WiMAX Scheduling Algorithms" WCECS 2009, October 20-22, 2009, San Francisco, USA

[11] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, "Quality of service support in IEEE 802.16 networks," IEEE Network, vol. 20,pp. 50-55, April 2006.

[12] N. Ruangchaijatupon and Y. Ji, '"Simple Proportional Fairness Scheduling for OFDMA Frame-Based Wireless Systems," in Proc. IEEE Wireless Communication and Networking Conf., Las Vegas, NV, 2008, pp. 1593-120

[13] www.opnet.com