An Efficient Quality of Service Scheduling Strategy for IEEE 802.16 Broadband Wireless Access Systems

Prasun Chowdhury[1] and Iti Saha Misra[2]

Electronics and Telecommunication Engineering
Jadavpur University, Kolkata, India
prasun.jucal@gmail.com, itisahamisra@yahoo.co.in

Abstract. In this paper, an efficient QoS scheduling strategy is proposed for IEEE 802.16 BWA systems that satisfies both throughput and delay guarantee to the applications such as video and audio streaming, online gaming, video conferencing, Voice over IP (VoIP), File Transfer Protocol (FTP) and Web Browsing. The proposed QoS scheduling strategy is compared with an existing QoS scheduling scheme proposed in literature in recent past. Simulation results show that the proposed scheduling architecture can provide a tight QoS guarantee for all types of traffic as defined in the standard hence maintain fairness and helps to eliminate starvation of lower priority class services.

Keywords: IEEE 802.16; MAC; QoS; Packet Scheduling.

1 Introduction

WiMAX (Worldwide Interoperability for Microwave Access), is a technology aimed at providing last-mile wireless broadband access at a cheaper cost. The "last mile" is the final leg of delivering connectivity from the service provider to the customer [1]. This leg is typically seen as an expensive undertaking because of the considerable costs of wires and cables. The core of WiMAX technology is specified by the IEEE 802.16 standard that provides specifications for the Medium Access Control (MAC) and Physical (PHY) layers. The term WiMAX was created by the WiMAX forum that promotes conformance and interoperability of the standard.

In WiMAX network such as, traffic from the Base Station (BS) to the Subscriber Stations (SSs) is classified as downlink traffic while that from the SSs to the BS is classified as uplink traffic. A scheduling algorithm implemented at the BS has to deal with both uplink and downlink traffic. In some cases, separate scheduling algorithms are implemented for the uplink and downlink traffic. Typically, a Call Admission Control (CAC) procedure is also implemented at the BS that ensures the load supplied by the SSs can be handled by the network [1-5]. A CAC algorithm will admit a SS into the network if it can ensure that the minimum Quality of Service (QoS) requirements of the SS can be satisfied and the QoS of existing SSs will not deteriorate. The performance of the scheduling algorithm for the uplink traffic strongly depends on the CAC algorithm.

Packet scheduling [5-14] is the process of resolving contention for shared re-sources in a network. The process involves allocating bandwidth among the users and determining their transmission order. Scheduling algorithms for a particular network need to be selected based on the type of users in the network and their QoS require-ments. QoS requirements vary depending on the type of application/user. For real-time applications such as video conferencing, voice chat and audio/video streaming, delay and delay jitter are the most important QoS requirements. Delay jitter is the inter-packet arrival time at the receiver and is required to be reasonably stable by the real-time applications. On the other hand, for non-real time applications such as file transfer (FTP), throughput is the most important QoS requirement. Some applications, such as web-browsing and email do not have any QoS requirements. In a network, different types of applications, with diverse QoS requirements, can co-exist. A sched-uling algorithm's task in a multi-class network is to categorize the users into one of the pre-defined classes. Each user is assigned a priority, taking into account its QoS requirements. Subsequently, bandwidth is allocated according to the priority of the users as well as ensuring that fairness between the users is maintained.

Packet scheduling algorithms are implemented at both the BS and SSs. A scheduling algorithm at the SS is required to distribute the bandwidth allocation from the BS among its connections. A scheduling algorithm at the SS is not needed if the BS grants bandwidth to each connection of the SS separately i.e. the Grant per Connection (GPC) procedure is followed. If the Grant per Subscriber Station (GPSS) procedure is followed, the scheduling algorithm at the SS needs to decide on the allocation of bandwidth among its connections. The scheduling algorithm implemented at the SS can be different than that at the BS [8].

The focus of our work is on scheduling algorithms for the uplink traffic in WiMAX i.e. traffic from the SSs to the BS. A scheduling algorithm for the uplink traffic is faced with challenges not faced by an algorithm for the downlink traffic. An uplink scheduling algorithm does not have all the information about the SSs such as the queue size. An uplink algorithm at the BS has to coordinate its decision with all the SSs where as a downlink algorithm is only concerned in communicating the decision locally to the BS.

K. Wongthavarawat *et al.* propose a hybrid scheduling algorithm in [1] that combines Earliest Deadline First (EDF), Weighted Fair queuing (WFQ) and First in First out (FIFO) scheduling algorithms. The overall allocation of bandwidth is done in a strict priority manner i.e. all the higher priority SSs are allocated bandwidth until they do not have any packets to send. The EDF scheduling algorithm is used for SSs of the rtPS class, WFQ is used for SSs of the Non-Real Time Polling Service (nrtPS) class and FIFO for SSs of the Best Effort (BE) class. Besides the scheduling algorithm, an admission control procedure and a traffic policing mechanism are also proposed. All these components together constitute the proposed QoS architecture. A drawback of this algorithm is that lower priority SSs will essentially starve in the presence of a large number of higher priority SSs due to the strict priority overall bandwidth allocation.

J. Lin *et al.* [8] propose architecture called Multi-class Uplink Fair Scheduling Structure (MUFSS) to satisfy throughput and delay requirements of the multi-class traffic in WiMAX. The proposed scheduling discipline at the BS is Modified Weighted Round Robin (MWRR), although details of the modifications to the

Weighted Round Robin (WRR) discipline are not provided by the authors. The model is based on Grant per Subscriber Station (GPSS) bandwidth grant mode and thus schedulers are implemented at the SSs to distribute the bandwidth granted among their connections. At the SS, Modified WFQ (MWFQ) is used for Unsolicited Grant Service (UGS) and Real-Time Polling Service (rtPS) connections, MWRR is used for nrtPS connections and FIFO is used for BE connections.

K. Vinay *et al.* [9] propose a hybrid scheme that uses EDF for SSs of the rtPS class and WFQ for SSs of the nrtPS and BE classes. This algorithm differs from the one in [1] in a couple of ways. First, the WFQ algorithm is used for SSs of both nrtPS and BE classes. Secondly, the overall bandwidth allocation is not done in a strict priority manner. Although the details of overall bandwidth allocation are not specified, it is briefly mentioned that the bandwidth is allocated among the classes in a fair manner. Since SSs of the BE class do not have any QoS requirements, using a computationally complex algorithm such as WFQ for them is not needed. Here Author made the comparative study of the scheduling algorithms implemented in GPSS and GPC and found that GPSS gives better end-to-end delay.

M.Settembre *et al.* [10] propose a hybrid scheduling algorithm that uses WRR and Round Robin (RR) algorithms with a strict priority mechanism for overall bandwidth allocation. In the initial portion of the algorithm, bandwidth is allocated on a strict priority basis to SSs of the rtPS and nrtPS classes only. After that the WRR algorithm is used to allocate bandwidth among SSs of rtPS and nrtPS classes until they are satisfied. If any bandwidth remains, it is distributed among the SSs of the BE class using the RR algorithm. This algorithm will starve lower priority SSs in the presence of a large number of higher priority SSs. The algorithm can also result in low fairness among SSs as it selects SSs with the most robust burst profiles first.

J. SUN *et al.* [11] proposed that the scheduler inside the BS may have only limited or even outdated information about the current state of each uplink connection due to the large Round Trip Delay (RTD) and possible collision occurred in the uplink channel transmission. So there is a need of an additional scheduler in each SS to reassign the received transmission opportunities among different connections. Since the uplink traffic is generated at SS, the distributed scheduler is able to arrange the transmission based on the up-to-date information and then provide QoS guarantee for its connections. But here the proposed algorithm is suffered by a problem called as *starvation of lower priority class services*.

In the proposed method the uplink traffic is scheduled based on current queue information at SS similarly in the way proposed in [11]. But a different hybrid algorithm has been implemented at the SS scheduler which helps to eliminate the starvation of lower priority class services. In the proposed method the BS scheduler can guarantee the minimum bandwidth for each service flow and ensure fairness and QoS in distributing excess bandwidth among all connections. At the same time, the scheduler in SS can provide differentiated and flexible QoS support for all of the four scheduling service types. In this paper EDF algorithm is applied for rtPS class of services and Deficit Fair Priority Queue (DFPQ) algorithm found in literature [12],[13] is applied for nrtPS and BE class of services. It can both reduce the delay of real-time applications and guarantee the throughput of non-real-time applications such as nrtPS and BE.

The rest of the paper is organized as follows; The QoS related features of IEEE 802.16 standard are discussed in section 2. Then the proposed scheduling algorithms are introduced in section 3. Section 4 provides simulation model and results analysis. Finally, this paper is ended up with the conclusions drawn in section 5.

2   QoS Features of IEEE 802.16

The first version, known as 802.16, was completed in October 2001. It specified a Single Carrier (SC) air interface for fixed point-to-multipoint (PMP) BWA systems operating between 10-66 GHz. The second amendment, 802.16a, was published in January 2003. It extends the physical environment towards lower frequency bands below 11 GHz. The next approved version is 802.16d, which is published in June 2004 and also known as FIXED WiMAX (802.16-2004). It incorporates all the previous versions to provide fixed BWA. In 2005, IEEE is undertaking the standardization of 802.16e, which is expected to support full mobility up to 70-80m/s [6]. Four service types are defined in IEEE 802.16d-2004 (Fixed) standard, which includes UGS (Unsolicited Grant Service), rtPS (Real-time Polling Service), nrtPS (Non Real-time Polling Service), and BE (Best Effort). The UGS is designed to support real-time service flow that generates fixed-size data periodically, such as T1/E1, VoIP without silence suppression [1], [6]. The rtPS is designed to support real-time service flow that generates variable size data, such as video streaming services while the nrtPS deals with FTP in similar manner. The BE perform tasks related to e-mail and web browsing. The guaranteed delay aspect is also taken care in video streaming and VoIP.

Since IEEE 802.16 MAC protocol is connection oriented, the application first establishes the connection with the BS as well as the associated service flow (UGS, rtPS, nrtPS or BE). BS will assign the connection with a unique connection ID (CID) [1], [6]. All packets from the application layer in the SS are classified by the connection classifier based on CID and are forwarded to the appropriate queue. So the scheduler inside the BS has outdated information about the current state of each uplink connection due to the large Round Trip Delay (RTD) and possible collision occurred in the uplink channel transmission [14].

3   Proposed Scheduling Algorithm

In the proposed method, the uplink bandwidth allocation at BS is done based on the per connection requests from SSs. Because a SS may have multiple connections at the same time, the bandwidth request messages should report the bandwidth requirement of each connection in SS. After that the allocated bandwidth per connection is pooled together and granted to each SS. Then SS re-distribute the received transmission opportunities among its connections according to their QoS requirement. Therefore an additional scheduler is needed in each SS to reassign the received transmission opportunities among different connections. Since the uplink traffic is generated at SS, the distributed scheduler is able to arrange the transmission based on the up-to-date information and then provide tight QoS guarantee for its connections.

Since the BS scheduler has limited information on the traffic generated at SS, the computing of bandwidth allocation should just consider the bandwidth request and reservation for each connection.

Let $BWMIN_i$ denote the minimum reserved bandwidth for connection i, and $BWREQ_i$ represent the bandwidth currently demanded by the connection i. Since the connection will never get more resources than it has requested, the bandwidth actually allocated during this phase is

$$BWALLOCATE_i = min \{BWMIN_i, BWREQ_i\} \tag{1}$$

For rtPS and nrtPS, $BWMIN_i$ is specified by the QoS parameter termed Minimum Reserved Traffic Rate. Clearly, to guarantee the contracted bandwidth, the sum of minimum reserved bandwidth for all the connections should not exceed the available bandwidth B. After each connection gets its guaranteed bandwidth, if there is still excess uplink bandwidth remained, BS scheduler should distribute the residual bandwidth in proportion to the pre-assigned connection weight. The algorithm in this phase can be described as:

```
BWREMAIN = B - ∑ BWALLOCATEᵢ;
               i
While BWREMAIN > 0
{
     If (BWALLOCATEᵢ < BWREQᵢ)
     {
                                  n
          BWADDᵢ = BWREMAIN*Wᵢ / ∑Wₖ;
                                 k=1
          BWALLOCATEᵢ = BWALLOCATEᵢ + BWADDᵢ;
          BWREMAIN = BWREMAIN - BWADDᵢ;
     }
}
```

Where BWREMAIN is the remaining bandwidth, $BWADD_i$ is the amount of excess bandwidth allocated to connection queue i and W_i is the weight of connection queue i. Now the allocated bandwidth per connection is pooled together and granted to each SS.

SS scheduler will select the packet to be transmitted from the highest priority queue. The priority of the queue is maintained in the following way UGS > rtPS > nrtPS > BE.

- Scheduling algorithm for UGS queues
UGS generate fixed size data packets on a periodic basis. This service has a critical delay and delay jitter requirement. So SS scheduler will firstly guarantee the bandwidth for UGS queues.

- Scheduling algorithm for rtPS queues
For rtPS service, Each packet entering the rtPS queues should be marked with a delivery deadline equal to t + tolerated delay, where t is the arrival time and tolerated delay is

the Maximum Latency for such a service flow. The packet with smaller deadline will be transmitted earlier. This greatly reduces the end-to-end delay of rtPS service.

- Scheduling algorithm for nrtPS and BE queues
For nrtPS and BE services Deficit Fair Priority Queue (DFPQ) algorithm found in literature [12], [13] is employed. DFPQ is almost similar to Deficit Round Robin (DRR) algorithm. This algorithm has been applied because of the following reasons.

1. The algorithm is mostly suited for datagram networks where packet sizes vary.
2. Since this algorithm requires accurate knowledge of packet size, it is suitable for the uplink traffic at SS scheduler.
3. The algorithm is flexible enough as it allows provision of quanta of different sizes depending   on the QoS requirements of the SSs. With this algorithm employed, SS scheduler can guarantee the minimum bandwidth for every non real time services such as nrtPS and BE connection and hence maintain an acceptable throughput. *Thereby eliminate starvation of lower priority service classes.*

In each service round, the nrtPS queue will be served first until its assigned bandwidth is deficit. If the assigned bandwidth is deficit for the nrtPS queue, the BE service flow queue will have chance to be served. Similar to [12], in DFPQ algorithm, a Quantum Q is assigned to each queue *i*. The quantum of a queue *i (Q[i])* represent the maximum number of bits that can be serviced in the first round. The scheduler visits each nonempty queue and determines the number of bandwidth requests in this queue. If there are more packets in the queue *i* after servicing *Q[i]* bits, the remaining amount of bits is stored in a queue state variable called *Deficit Counter (DC[i])* and the scheduler moves on to serve the next non-empty queue. In subsequent rounds, the amount of bandwidth usable by this flow is the sum of *DC[i]* in the previous round added to *Q[i]*. The *Q[i]* is the Maximum Sustained traffic rate (r̄max) of a certain service flow. In case *rmax* = 0 (BE service flow), *rmin* is used instead. As a result of using the quantum variable, connections with larger quantum are serviced more.

DFPQ algorithm is shown below. Here *Ltotal* is defined to be the remaining total capacity of the frame after servicing UGS and rtPS queues and *La* to be its remaining capacity. p(i,k) denotes k[th] packet of ith connection, i belongs to nrtPS and BE connections. The algorithm can be described as follows,

Each service queue i of nrtPs and BE is initialized with

```
La = Ltotal;
DC[i] = Q[i];

While La > 0
{
        While DC[i]>0 and p (i, k) is available
        {
            DC[i] = DC[i] - p(i,k);
            La =La - p(i,k);
```

```
        k++;
    }
    DC[i] = DC[i] + Q[i];
    i = (i % n) + 1;
}
Go to UGS scheduling for next frame
```

4   Simulation Model and Result Analysis

To evaluate the effectiveness and efficiency of the proposed scheduler, the IEEE
802.16 MAC layer protocol is analyzed using MATLAB. A number of simulations
are conducted in this section. At first, the simulation environment and parameters are
described. And then the simulation results will be presented for discussion.

4.1   Simulation Environment and Parameters

As mentioned before, basically PMP MAC operation is focused in this article. A
TDD-OFDM system is used in our simulation with the MAC layer configuration
parameters as shown in table 1 and the network is configured as consists of one BS
and multiple SSs as shown in fig 1.
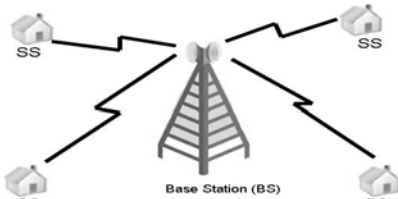
Fig. 1. Proposed model architecture

Table 1. MAC layer configuration parameters

Bandwidth	4.3MHz
Frame duration	10 ms
Duplex	TDD
Delay	20ms

4.2   Simulation Result and Discussions

The simulated 802.16 network consists of one BS and four SSs with different traffic
patterns. The first SS is configured with all types of traffic flows nominated as UGS-
_1, rtPS_1, nrtPS_1 and BE_1, the second SS has UGS_2, rtPS_2, nrtPS_2 and BE_2,
the third SS has UGS_3, rtPS_3, nrtPS_3 and BE_3, the fourth SS runs UGS_4,
rtPS_4, nrtPS_4 and BE_4. Two scenarios - with and without SS scheduler are simu-
lated to study the effect of SS scheduler.

   Our proposed method is compared with the proposed method in [11]. Here, "with-
out SS scheduler" and "BS-(service class)" means that BS scheduler will designate
bandwidth to individual connection. On the other hand, "with ss-scheduler1" and
"SS1-(service class)" means that SS scheduler will designate bandwidth to individual
connection in our proposed method. Again "with ss-scheduler2" and "SS2-(service
class)" means that SS scheduler will designate bandwidth to individual connection
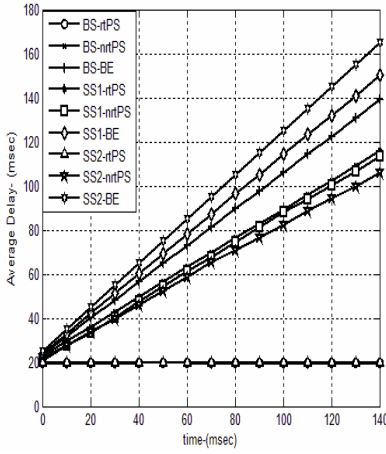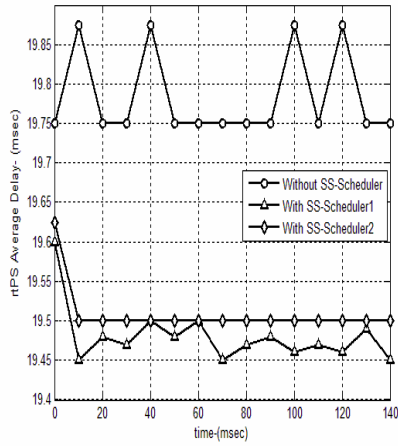
Fig. 2. Service delay comparison
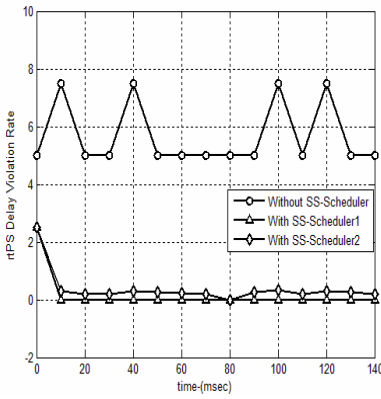

Fig. 3. Service delay comparison of rtPS
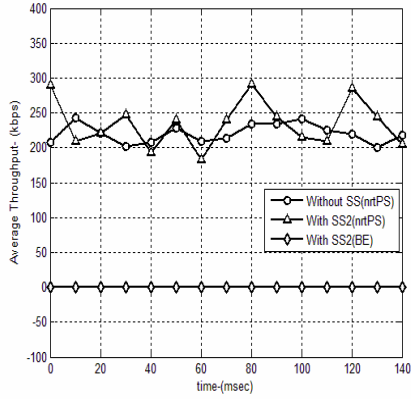

Fig. 4. % of packet drop comparison

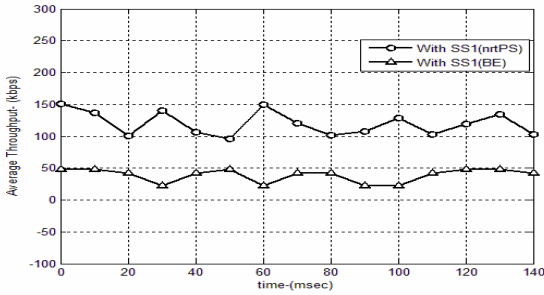**Fig. 5.** Throughput comparison proposed in [11]

Fig. 6. Throughput comparison in our proposed strategy

proposed in [11]. Here service class refers to UGS, rtPS, nrtPS and BE**.** As UGS generate fixed size data packets on a periodic basis so delay is negligible and throughput is constant and hence it is not shown in our simulation result. Fig. 2 displays the end-to-end delay of different services with and without SS scheduler. The curves show that after SS scheduling, low priority service suffered longer delay. From rtPS, nrtPS to BE, the end-to-end delay increased with the service priority decreased. The fundamental requirement of QoS scheduling for IEEE 802.16 systems is achieved.

To further demonstrate this benefit, we simulated the rtPS performance under the same number of background SS in Fig. 3. From fig.4 we can see that the SS scheduler can effectively reduce the QoS violation rate of rtPS service flow. Here, the QoS violation rate is defined as the amount of packets whose delay is larger than the Maximum Latency to the total amount of packets that have been received from network interface.

By introducing DFPQ algorithm for nrtPS and BE services on priority order, SS scheduler can guarantee the throughput of nrtPS service as well as BE service. Though BE services do not require any QoS but in our proposed algorithm fairness is maintained for all type of service classes. But this fairness is not maintained in [11] where we got throughput of BE service is zero shown in fig.5. But our simulation result shows that SS scheduler guarantees the throughput of nrtPS service as well as BE service shown in Fig.6 where throughput of BE service is greater than zero throughout simulation time, hence maintain fairness of all the services and eliminate the problem of starvation of lower priority class services.

5   Conclusion

We proposed an efficient packet scheduling scheme for IEEE 802.16 WiMAX to satisfy both delay and throughput guarantees for the admitted connections. An architecture model was developed to demonstrate the performance of the proposed scheme. The simulation results show that the proposed scheme is the best choice for QoS scheduling in WiMAX in terms of delay and throughput of all connections of the system compared to schemes proposed in [11]. Simulation results prove that the BS scheduler can guarantee the minimum bandwidth for each service flow and ensure fairness and QoS in distributing excess bandwidth among all connections. At the same time, the scheduler in SS can provide differentiated and flexible QoS support for all of the four scheduling service types. It can both reduce the delay of real-time applications and guarantee the throughput of non-real-time applications. Thereby eliminate starvation problem of lower priority class services. Therefore, the proposed QoS scheduling architecture can provide good optimized QoS guarantees for all types of traffic classes as defined in the scheduler standard.

References

1. Wongthavarawat, K., Ganz, A.: Packet scheduling for QoS support in IEEE 802.16 broad-band wireless access systems. International Journal of Communication Systems 16(1), 81–96 (2003)
2. Wang, H., Li, W., Agrawal, D.P.: Dynamic admission control and QoS for IEEE 802.16 Wireless MAN. In: Proc. of Wireless Telecommunications Symposium (WTS 2005), April 6-7, pp. 60–66 (2005)
3. Niyato, D., Hossain, E.: Connection Admission Control Algorithms for OFDM Wireless Networks. In: Proc. of IEEE Globecom, pp. 2455–2459 (2005)
4. Suresh, K., Misra, I.S., Saha (Roy), K.: Bandwidth and Delay Guaranteed Call Admission Control Scheme for QOS Provisioning in IEEE 802.16e Mobile WiMAX. In: Proceedings of IEEE GLOBECOM (December 2008)
5. Jiang, C.-H., Tsai, T.-C.: CAC and Packet Scheduling Using Token bucket for IEEE 802.16 Networks. In: 3rd IEEE, Consumer Communications and Networking Conf., CCNC 2006, January 8-10, vol. 1, pp. 183–187 (2006)
6. Andrews, J.G., Ghosh, A., Muhamed, R.: Fundamentals of WiMAX-Understanding Broadband Wireless Networking. Pearson Education, London (March 2007)
7. Chowdhury, P., Misra, I.S.: A Comparative Study of Different Packet Scheduling Algorithms with Varied Network Service Load in IEEE 802.16 Broadband Wireless Access Systems. In: IEEE Proc. Int. Conf. Advanced Computing & Communications, Bangalore (December 2009)
8. Lin, J., Sirisena, H.: Quality of Service Scheduling in IEEE 802.16 Broadband Wireless Networks. In: Proceedings of First International Conference on Industrial and Information Systems, August 2006, pp. 396–401 (2006)
9. Vinay, K., Sreenivasulu, N., Jayaram, D., Das, D.: Performance evaluation of end-to-end delay by hybrid scheduling algorithm for QoS in IEEE 802.16 network. In: Proceedings of International Conference on Wireless and Optical Communication Networks, 5p. (April 2006)
10. Settembre, M., Puleri, M., Garritano, S., Testa, P., Albanese, R., Mancini, M., Lo Curto, V.: Performance analysis of an efficient packet-based IEEE 802.16 MAC supporting adaptive modulation and coding. In: Proceedings of International Symposium on Computer Networks, pp. 11–16 (June 2006)
11. Sun, J., Yao, Y., Zhu, H.: Quality of Service Scheduling For 802.16 Broadband Wireless Access System. In: Advanced system technology telecom lab, China, Beijing, IEEE, Los Alamitos (2006)
12. Shreedhar, M., Varghese, G.: Efficient Fair Queuing using Deficit Round Robin. IEEE/ACM Transactions on Networking 1(3), 375–385 (1996)
13. Safa, H., Artail, H., Karam, M., Soudah, R., Khyat, S.: New Scheduling Architecture for IEEE 802.16 Wireless Metropoliton Area Network, American university of Beirut, Lebanon. IEEE, Los Alamitos (2007)
14. Chu, C., Wang, D., Mei, S.: A QoS architecture for the MAC protocol of IEEE 802.16 BWA system. IEEE Communications, 435–439 (July 2002)