CAC and Uplink Scheduling Algorithms in WiMAX Networks

Eduardo Castilho Rosa & Paulo Roberto Guardieiro

***Abstract*—The IEEE 802.16 standard meets the need to provide wireless broadband connectivity for both mobile and fixed users. Because the standard does not specify the implementation of mechanisms for quality of service (QoS) provisioning, in this paper we propose a scheme for QoS provision which consists of an uplink scheduling mechanism in the base station (BS) and a connection admission control (CAC) policy based on the prediction of network delay. The algorithm scheduling is designed to support four types of service flows (UGS, rtPS, nrtPS and BE) and the admission control scheme makes a prediction on network delay for decision making. The prediction delay is calculated according to the queue current size in the subscribers stations (SSs), which are sent to the BS periodically by means of bandwidth request mechanism. We analyzed the proposed scheme using modeling and simulation and our simulation results were sufficient to provide QoS guarantees.**

***Index Terms*—CAC, Scheduling, Quality of Service, IEEE 802.16, WiMAX, Delay Prediction**

I. Introduction

With the emergence and growth of applications with heterogeneous traffic (voice, video and data), the IEEE 802.16 standard [1], also known by the acronym WiMAX (Worldwide Interoperability for Microwave Access), is becoming an attractive option for wireless broadband access to last mile. This is mainly because these networks offer a good cost-benefit to the end user, i.e., high capacity data transmission at a relatively low cost of deployment.

When compared to traditional wired access technologies, the IEEE 802.16 standard has the advantage of allowing rapid delivery of services in areas of difficult access. Thus, the WiMAX networks allows us to accelerate the introduction of broadband wireless technology in the market, as well as increase performance and reliability of services offered by service providers [2].

The main feature incorporated by IEEE 802.16 standard, which makes it a candidate to represent fourth-generation (4G) wireless communication systems, is the differentiated treatment of traffic generated by applications, essential to QoS provisioning. Furthermore, the IEEE 802.16 standard requires scheduling policies, traffic policing and admission control schemes for complete the QoS provisioning architecture. However, in order to intensify the competition among network equipment manufacturers, these mechanisms are not defined by the standard. Thus, the standard enables the proposal of solutions that meet the QoS requirements while maintaining a diversification of products, that allows the choice based on the required performance.

Since the IEEE 802.16 standard does not define the policies for admission control and packet scheduling, in this paper we propose a QoS provisioning mechanism which consists of a new scheduling algorithm for uplink traffic and a dynamic admission control scheme for IEEE 802.16 networks. The uplink scheduling algorithm works in conjunction with a bandwidth reservation mechanism and prioritize the connections that have a large amount of packets in their queues. The admission control policy, on the other hand, ensures that the entry of a new connection in the network does not affect the QoS requirements of existing connections. This algorithm is based on a delay prediction scheme and uses the buffer size information in the subscriber station (SS), which are sent to BS periodically, through of bandwidth request mechanism. Simulations were conducted to evaluate the experiments using different applications classes.

The remainder of this paper is organized as follows: In Section II, we give an introduction of IEEE 802.16 standard. In Section III, we describe our proposed scheme and the Section IV shows the related works. In the Section V we present our simulation scenario. Section VI provides an analysis of the simulation results. Finally, the conclusions are presented in Section VII.

II. The IEEE 802.16 Standard

The IEEE 802.16 standard is based on OSI (Open Systems Interconnection) model and specifies the physical and MAC (Media Access Control) layers in order to enable the wireless broadband internet access. The MAC layer is situated just above the physical layer and your main task is to provide an interface between the upper layers (or other packet-based networks). The protocols that operate within the MAC layer are responsible for performing the main functions of the IEEE 802.16 standard, including the mechanisms for QoS provisioning and mobility management. The physical layer, on the other hand, has the function of transmitting the bits over the wireless channel by means of a modulation and codification scheme. It operates at 10-66 GHz for Line-of-Sight (LOS) environments and 2-11 GHz for Non Line-of-Sight (NLOS) environments with data rates of 32-130 Mbps, according to the available bandwidth in the channel [3].

The standard also defines two architectures related to communication mode: point-to-multipoint mode (PMP) and mesh mode. In PMP mode, every SSs communicates directly with

Manuscrito recebido em 30 de novembro de 2010; revisado em 1 de abril de 2011.

E. C. Rosa (eduardocastilho@iftm.edu.br) is with Instituto Federal do Triângulo Mineiro - (IFTM) - Av. Córrego do Pirapitinga, s/n - Ituiutaba - MG - Brazil - 38.300-000

P. R. Guardieiro (prguardieiro@ufu.br) is with Universidade Federal de Uberlândia - (UFU) - Av. João Naves de Ávila, 2121 - Uberlândia - MG - Brazil - 38.400-902

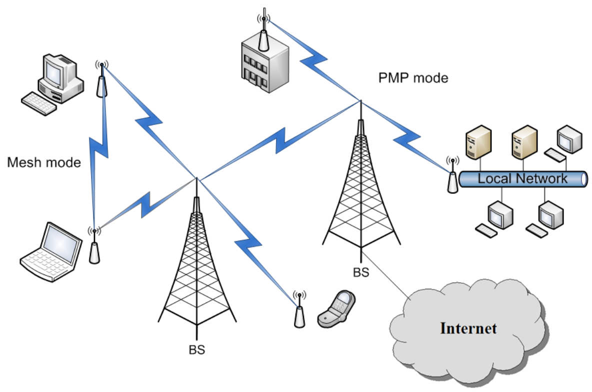

Fig. 1. Operation modes in the WiMAX network.

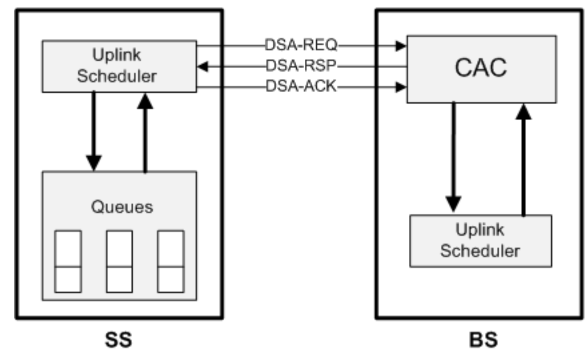

Fig. 2. Connection establishment in IEEE 802.16 standard.

the BS (SS-BS-SS) forming a star topology network. This architecture facilitates the network design by centralizing the communication management within BS. In the mesh mode, the SSs can exchange information without interference from the BS (SS-SS). However, the complexity in this operation mode is greater because the SS has at least an additional control module to manage the communication in BS. Figure 1 shows the operation modes in the WiMAX network.

The control over the wireless link sharing is performed at the physical layer by means of time division duplexing (TDD) and the frequency division duplexing (FDD). In TDD mode, a transmission frame is divided in time-domain into downlink and uplink subframes. In the downlink subframe, data is broadcasted to every SS using the entire frequency spectrum available. The transmitted data in this direction are distributed to each SS using a downlink map (DL-MAP), which contains the timestamp that each station must receive the data. In addition, the IEEE 802.16 standard also defines an uplink map (UL-MAP), which contains the timestamp that each station must transfer the data. In FDD mode, the frequency spectrum is divided into two parts, one for the downlink and one for uplink, except that data transmission can be performed simultaneously.

A. *QoS Provisioning in WiMAX Networks*

For the purpose of to support a wide variety of applications, the IEEE 802.16 standard was designed to provide QoS for user applications. For this, the standard is connection-oriented in the MAC layer. Each connection is identified by an unique connection identifier (CID), which is associated with a service flow characterized by a set of QoS parameters, e.g, tolerable delay and minimum/maximum traffic rate. The connection establishment is performed by using a three-way handshake mechanism, which is composed by DSA-REQ, DSA-RSP and DSA-ACK messages, as illustrated in Figure 2.

The main element that acts in the setting up a new connection is the CAC. It will receive a set of QoS parameters within the DSA-REQ message and will decide whether to accept the connection or not, according to the allocated bandwidth of admitted connections and the available bandwidth in the channel. This decision is important because it ensures that the admission of a new connection will not affect the QoS guarantee of connections already in service.

The IEEE 802.16 standard, as mentioned before, differentiates the user data in order to provide QoS assurances. For this, the standard specifies four service classes, namely: Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non Real-time Polling Service (nrtPS) and Best Effort (BE). The packets scheduler, which determines the order of their transmission, will work according to different priority levels: UGS > rtPS > nrtPS > BE. These classes are specified as follows:

- UGS - The UGS is designed to support real-time service flows that generate fixed-size data packets in periodic intervals, such as T1/E1 and Voice over IP without silence suppression. For this service BS offers fixed size unsolicited data grants, that is, the transmission opportunities on periodic intervals without any explicit request from the SS. This eliminates the overhead and latency of bandwidth requests. In the UGS service, the BS offers a fixed size grants on a real time periodic basis. For UGS SS is prohibited from using any contention request opportunities. As SS does not have to make any explicit bandwidth requests this method eliminates the overhead and latency of SS requests and ensure that grants are available to meet the flow's real-time needs [13].
- rtPS - The rtPS is designed to support real-time uplink service flows that generate transport variable size data packets on a periodic basis, such as moving pictures experts group (MPEG) video. E1/T1 type data services are also supported by rtPS. For fixed operators, rtPS guarantees E1/T1 data rates to allowing customers to burst higher when extra capacity is available on the network. rtPS service offers real-time, periodic, unicast request opportunities, which meet the flow's real-time needs and allow the SS to specify the size of the desired grant. This service requires more request overhead than UGS, but supports variable grant sizes for optimum data transport efficiency [13].
- nrtPS - The nrtPS is designed to support delay-tolerant data streams consisting of variable-sized data packets for which a minimum data rate is required, such as FTP. This service offers unicast polls on a regular basis to ensure assures that the service flow receives request

opportunities even during network congestion. The BS grants unicast polls to nrtPS connections on an interval of 1 s or less. The non real-time polling service is almost identical to the real time polling service. The difference is that for nrtPS, connections may utilize random access transmit opportunities for sending the bandwidth requests. The mandatory QoS service flow parameters for this scheduling service are Minimum Reserved Traffic Rate, Maximum Sustained Traffic Rate, Traffic Priority, and Request/Transmission Policy [13].

- BE - Unlike nrtPS, the Best Effort services are designed for applications that do not have any specific delay requirements. For BE the QoS parameters are chosen such that they provide scheduling service to support data streams for which no minimum resources allocation are granted. Therefore may be handled on a space-available basis. For BE services there is no QoS guarantee, like the email or the short length FTP. The only difference between nrtPS and BE services is that nrtPS connections are reserved a minimum amount of bandwidth using the minimum reserved traffic rate parameter. Both the nrtPS and BE services request bandwidths by either responding to the broadcast polls from the BS or piggybacking a bandwidth request [13].

III. The Proposed QoS Mechanism

The QoS provisioning in the IEEE 802.16 standard is performed through of packets scheduling mechanism and CAC policies (a comprehensive survey of CAC policies can be obtained in [8]) which are not specified by the standard, according to Figure 3. Thus, in this paper we propose an uplink scheduling algorithm for the UGS, rtPS, nrtPS and BE service flows in the BS and an CAC policy for rtPS flows based on delay prediction. This mechanism was introduced in [11] and in this paper we include some additional results.

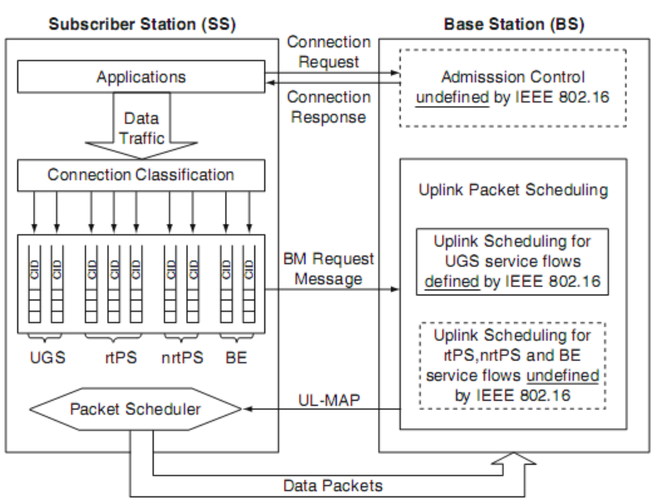

Fig. 3. QoS architecture of IEEE 802.16 [13].

A. Proposed Scheduling Mechanism

The proposed scheduling mechanism is based on a bandwidth reservation mechanism with static portions to logically separate the real-time flows (UGS and rtPS), non real-time flows with the guaranteed minimum bandwidth requirement (nrtPS) and best effort traffic (BE). This bandwidth reservation mechanism has been created in order to avoid starvation of low priority flows (bandwidth starvation), such as nrtPS and BE classes, when the network is operating under a high traffic load.

The proposed bandwidth reservation mechanism allocates a fixed amount of bandwidth for the connections belonging to UGS and rtPS service classes (W bps), which is allocated on demand for flows of these classes, according to the priority level UGS $>$ rtPS. The W portion is used primarily for the real-time service flows, but to avoid waste of bandwidth, W may be used to nrtPS and BE service flows if there are no UGS and rtPS flows in the network or the total bandwidth required for these connections is less than W. The second portion is reserved for the nrtPS connections (T bps) and, just as W, T can also be used for other service flows if there are no nrtPS connections or if the nrtPS requested bandwidth is less than T. Finally, a relatively small portion is intended to meet the BE service flows (R bps), only to avoid bandwidth starvation in this class. Figure 4 illustrates the proposed bandwidth reservation scheme.

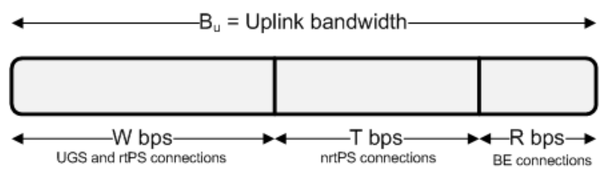

Fig. 4. Proposed bandwidth reservation scheme.

The proposed scheduling mechanism defines four queues in BS, one for each service class, and these are served in accordance to the priority levels specified for each service flow: UGS $>$ rtPS $>$ nrtPS $>$ BE. The UGS queue store the periodic grants for sending data while the rtPS, nrtPS and BE queues store the bandwidth requests messages. These queues are served in a similar mode to priority queuing (PQ) discipline. However, the queues are served preemptively based on W, T and R reserves.

In each scheduling cycle, the UGS connections are first served because of the required bandwidth is constant and guaranteed. In this case, the condition $\sum_{i=1}^{n_{ugs}} b_i \leq W$ (n_{ugs} = number of UGS connections and b_i = UGS rate) should be respected, otherwise a preemption is made and the next queue is served. Thereafter, the rtPS connections are served by ordering them according to queue size information. In this case, to avoid preemption in this queue, the following condition must be satisfied: $\sum_{i=1}^{n_{ugs}} b_i + \sum_{j=1}^{n_{rtPS}} \leq W$, where b_j^{req} is the value of the rtPS request related to j station and n_{rtPS} is the number of rtPS service flows in the network.

After serving the UGS and rtPS queues the nrtPS queue is served and the remaining bandwidth is divided between the nrtPS connections. In this way will be provided for each nrtPS flow the bandwidth $min(b_k^{req}, b_k^{max})$, where b_k^{req} is the value of the nrtPS request related to SS k and b_k^{max} is the average bandwidth resulting from the expression $(W +$

$T - (\sum_{i=1}^{n_{ugs}} b_i + \sum_{j=1}^{n_{rtPS}} b_j^{req}))/n_{nrtPS}$, where n_{nrtPS} is the number of SSs with nrtPS connections. Following this policy, the nrtPS connections will not suffer bandwidth starvation in the worst case (W is fully utilized), because T will be used only to nrtPS service flows. Finally, the BE connections are scheduled by distributing the remaining bandwidth $(W + T - (\sum_{i=1}^{n_{ugs}} b_i + \sum_{j=1}^{n_{rtPS}} b_j^{req} + \sum_{k=1}^{n_{nrtPS}} min(b_k^{req}, b_k^{max})))$ between each BE connection in the network. In the worst case, i.e., when both W and T are fully allocated to UGS, rtPS and nrtPS service flows, there still remains the R portion that will be allocated exclusively for the BE flows, avoiding the bandwidth starvation in this class. In the algorithm shown in Figure 5 we describe the proposed scheduling scheme for the UGS, rtPS, nrtPS and BE service classes.

B. Proposed Predictive CAC Algorithm

The proposed CAC mechanism is for the rtPS service class and works in conjunction with the proposed uplink scheduling algorithm. In the proposed CAC algorithm, upon receiving the DSA-REQ message, the BS makes an average delay prediction that a new connection can suffer in the network. If the predicted value is less than or equal to the threshold value, the connection is accepted. Otherwise, it is rejected. This condition is checked for each rtPS connection waiting to enter the network. The UGS, nrtPS and BE connections will be automatically accepted by the network, however, the amount of bandwidth allocated by the uplink scheduler for each service flow over time must satisfy W, T and R. Since average maximum delay is less than a threshold, the bandwidth is implicitly guaranteed for the accepted connections. The algorithm of Figure 6 shows the operations performed by the proposed CAC mechanism.

Upon admitting a new connection, the proposed scheme allocates a number of OFDM (Orthogonal Frequency Duplexing Modulation) symbols necessary to transmit all bandwidth requested, according to modulation coding scheme (MCS) employed. Thus, the calculation of the allocated bandwidth for all connections in the network varies stochastically over time, since it depends on the values of this request in the queue Q.

Considering there are, in a given time t, a maximum of K connections that can be scheduled in the current frame and there are N stations in the network, the average allocation for each station (b_i), which are stored in the queue Q_2 ($Q_{i,j}$ is the data grant or j request bandwidth in respect to queue i), is given by equations (1), (2) and (3):

$$b_i = \frac{\sum_j^K Q_{2,j}}{K} \quad (1)$$

subject to:

$$K * b_i \leq W - \sum_j^K Q_{1,j} \quad (2)$$

$$0 < K \leq N \quad (3)$$

Require: N_b = Total number of OFDM symbols in the uplink frame.
Ensure: UL-MAP = Uplink map.
for (j of 1 until $|Q_1|$) **do**
  $BW_{ugs} \leftarrow$ UGS symbols by frame..
  if $(BW_{ugs} + P > (N_b - (T + R)))$ **then**
    if $((N_b - (T + R)) < P)$ **then**
      Make up a preemption in UGS queue.
    end if
    $BW_{ugs} \leftarrow (N_b - (T + R))$;
  end if
  UL-MAP $\leftarrow$ Add BW_{ugs};
  $N_b \leftarrow N_b - BW_{ugs}$;
end for
$Q_2 \leftarrow$ Sorts rtPS queue.
for (j of 1 until $|Q_2|$) **do**
  $BW_j \leftarrow$ Symbols to transmit $Q_{2,j}$ bytes.
  if $(BW_j + P > (N_b - (T + R)))$ **then**
    if $((N_b - (T + R)) < P)$ **then**
      Make up a preemption in rtPS queue.
    end if
    $BW_j \leftarrow (N_b - (T + R))$;
  end if
  UL-MAP $\leftarrow$ Add BW_j;
  $N_b \leftarrow N_b - BW_j$;
end for
for (j of 1 until $|Q_3|$) **do**
  $BW_{reserved} \leftarrow (N_b - R)/|Q_3|$
  $BW_{allocated} \leftarrow min(BW_{requested}, BW_{reserved})$
  if $(BW_{allocated} + P > (N_b - R))$ **then**
    if $((N_b - R) < P)$ **then**
      Make up a preemption in nrtPS queue.
    end if
    $BW_{allocated} \leftarrow (N_b - R - P)$;
  end if
  UL-MAP $\leftarrow$ Add $BW_{allocated}$;
  $N_b \leftarrow N_b - BW_{allocated}$;
end for
$index\_ \leftarrow$ Last BE station served in the previous frame.
for (j of 1 until $|Q_4|$) **do**
  $i \leftarrow (j + index\_)$ mod $|Q_4|$;
  $BW_{be} \leftarrow$ Symbols to transmit i bytes.
  if $(BW_{be} + P > N_b)$ **then**
    if $((N_b < P)$ **then**
      Exit.
    end if
    $BW_{be} \leftarrow (N_b - P)$;
  end if
  UL-MAP $\leftarrow$ Add BW_{be};
  $N_b \leftarrow N_b - BW_{be}$;
end for
return UL-MAP;

Fig. 5. Proposed uplink scheduling algorithm.

In the proposed CAC mechanism, each request R_i that arrives from rtPS, nrtPS and BE stations, which reflects the applications queue length, is stored in three queues in the BS: one for rtPS, one for nrtPS and one for BE. The

Require: DSA-REQ message;
Ensure: DSA-RSP message;
1: $c \leftarrow$ DSA-REQ.$connection$;
2: **if** $(c.type = rtPS)$ **then**
3: $A_t \leftarrow$ Search the current predicted delay in the network;
4: **if** $(A_t < threshold)$ **then**
5: DSA-RSP.flag $\leftarrow 1$; //Accept c
6: **else**
7: DSA-RSP.flag $\leftarrow 0$; //Reject c
8: **end if**
9: **else**
10: DSA-RSP.flag $\leftarrow 1$; //Accept c.
11: **end if**
12: **return** DSA-RSP;

Fig. 6. Proposed CAC algorithm.

prediction module will access the contents of these queues and, in each time interval f (frame duration), will perform the prediction. Once the BS receives a new connection request in the DSA-REQ message, the current predicted value (line 3 from algorithm in Figure 6) will be used in the admission control process.

The equation (4) describes the dynamics of the average queue size (B_t) of the stations that can not have their connections scheduled in current frame, taking into account the time t, N connections and a maximum of K connections scheduled in current frame.

$$B_t = \begin{cases} 0, & \text{if N=K;} \\ \\ \dfrac{\sum_{i=1}^{N-K} R_i}{N-K}, & \text{if N>K;} \end{cases} \tag{4}$$

According to queue size estimation B_t, the prediction of network delay (A_t) can be calculated by equation (5):

$$A_t = \frac{B_t}{r} * T_s * f \tag{5}$$

where r is the modulation efficiency (bits/symbol), the OFDM symbol time is T_s (ms) and f is the frame duration (ms). The equation 5 shows that the predicted delay is directly proportional to B_t, since r, T_s and f is constant.

IV. Related Works

There are several works in the literature that discuss techniques for scheduling and CAC in the IEEE 802.16 standard. In this paper, we present four important works related to our proposed mechanisms.

In [4], a CAC scheme that uses bandwidth and delay information is presented. The bandwidth control is performed according to the fixed allocation criterion, reserving the minimum rate for each class. The maximum delay control, on the other hand, is performed according to numerical prediction of delay, where this value is compared to with the maximum delay requirement for decision-making in CAC. In this paper, we also make a delay prediction in the network, but it is made by means of real information in the queue of stations. This paper is a good contribution and serves as our primary reference.

In [5], a proposal of CAC and packet scheduling using token bucket is presented. In this paper, an estimation model of delay and packet loss is also provided using the token bucket with token rate (r_i) and bucket size (b) parameters. Our method also uses an estimation model of delay, but based on queue size in the SSs. This proposal also includes a discrete time Markov chain model to analyze the behavior of queues (infinite and finite queues).

In [6] and [7], a CAC scheme based on bandwidth reservation model is proposed. The decision to accept the connection is made according to fixed thresholds values for each class. However, the admission process takes into account only the bandwidth requirement. Our method, moreover, also takes into account the delay requirements.

V. Modeling and Simulation

Both proposed scheduling and CAC algorithms were implemented in NS-2 [9] with the WiMAX module developed by NIST (National Institute Standard Technologies) [12]. This simulator includes scheduling algorithms for UGS, rtPS (based on packets deadline) and nrtPS class, but does not include admission control algorithms. To implement the proposed algorithms, it was necessary to extend this WiMAX module in order to add connections over time.

The considered scenarios involve one BS and a variable number of SSs in the network at regular periods and random positions. The maximum distance allowed between a SS and the BS is 500 meters, which enables the use of a MCS more efficient [10].

Our simulation model considers one connection by station and the GPSS (Grant per Subscriber Stations) mode was used in granting the bandwidth. The main simulation and applications parameters are listed in Table I. These parameters were chosen because they were used in most studies in the literature.

TABLE I

Main simulation parameters.

PARAMETER	VALUE
Operating Frequency	3.5 GHz
Bandwidth	5 MHz
Duplexing	TDD
Antenna	Omnidirecional
Propagation Model	2-Ray Ground
Frame Duration	25 ms
Cyclic Prefix (CP)	0.25
Modulation	OFDM 64QAM 3/4
Uplink Rate	7.70 Mbps
UGS Traffic	CBR (Packet size=40 bytes; Interval=0.02s). Traffic Rate = 16 Kbps
rtPS Traffic	Video Streaming MPEG (Packet size=[200:1000]; Interval=0.01s). Average rate = 480 Kbps
nrtPS Traffic	FTP (Minimum rate = 160 Kbps; Maximum rate = 800 Kbps)
BE Traffic	Web Traffic (Average rate = 75 Kbps)
Delay Threshold	20 ms if service class = UGS; 200 ms if service class = rtPS.

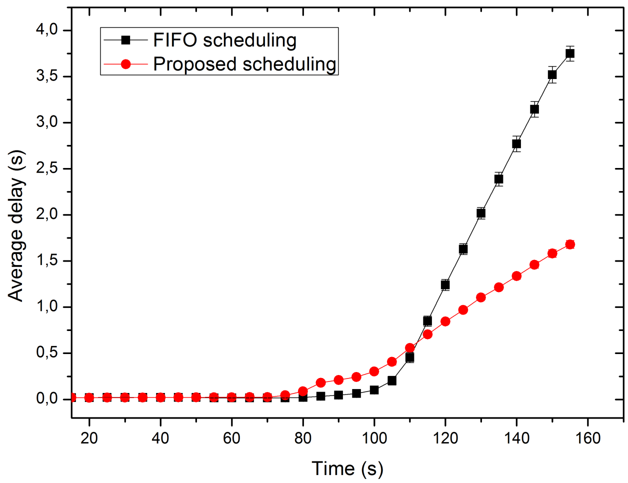

Fig. 7. Delay performance of the proposed scheduler comparison with FIFO scheduler.

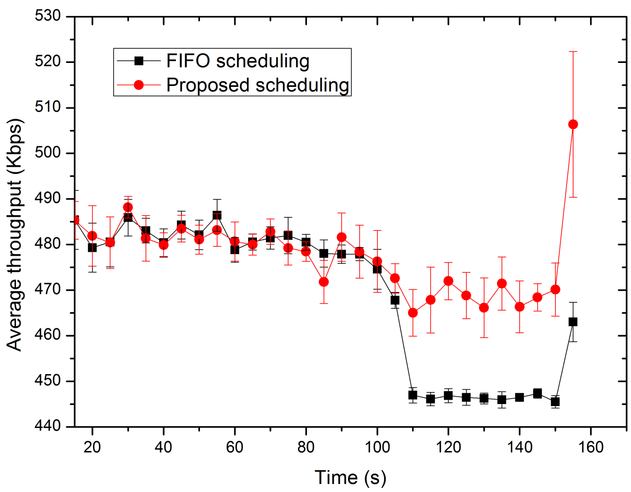

Fig. 8. Throughput performance of the proposed scheduler comparison with FIFO scheduler.

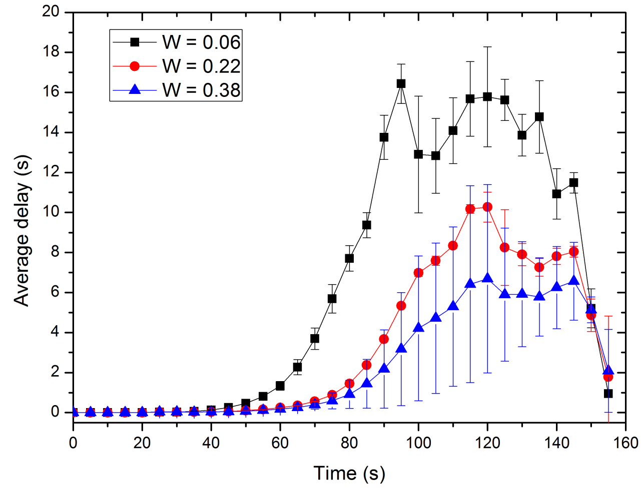

Fig. 9. Variation of W for rtPS flows in terms of average delay.

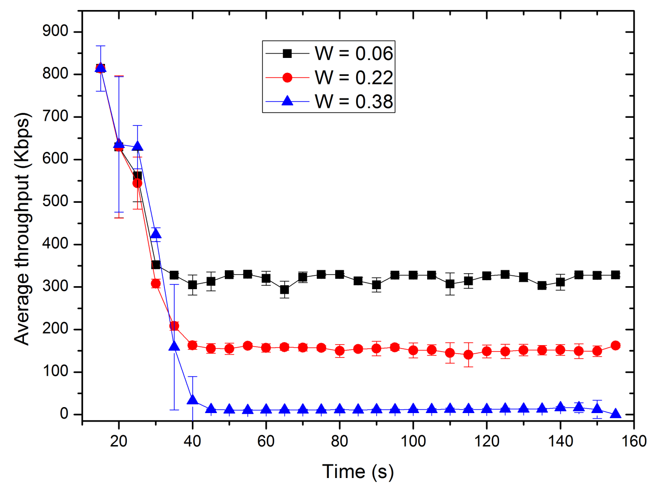

Fig. 10. Variation of W for rtPS flows in terms of average throughput.

VI. Simulation Results

The results showed in this section refer to five simulation rounds for each scenario in order to obtain a confidence interval of 95%. In all simulation experiments we considered the periodic arrival of rtPS stations on the network at fixed intervals of 5 seconds. After entering the network, each station begins to transmit data until the end of the simulation. The throughput (th) and average delay (d) are calculated at periodic intervals ($t, t{+}k$), taking $k = 5$, according to equations (6) and (7):

$$th^z_{t,t+k} = \frac{\sum_{i=1}^{T} size^{t,t+k}_{i,z}}{T} \tag{6}$$

$$d^z_{t,t+k} = \frac{\sum_{i=1}^{N} (\sum_{j=1}^{P_{cont,i}} (Rx^{t,t+k}_{j,z} - Tx^{t,t+k}_{j,z})/P_{cont,i})}{N} \tag{7}$$

where z = service class; T = Total number of packets received at $[t, t{+}k]$ interval; $size^{t,t+k}_{i,z}$ = ith packet size received at $[t, t+k]$ interval; $P_{cont,i}$ = amount of packets received refers to i-station; i; $Rx^{t,t+k}_{j,z}$ e $Tx^{t,t+k}_{j,z}$ = the receive and transmission time to j-packet, respectively; k = the sample interval (we uses 5 seconds); N = the amount of z-stations in the network at $[t,\ t+k]$ interval.

A. Performance Evaluation of the Proposed Uplink Scheduling Algorithm

1) Experiment 1: The first experiment was modeled to analyze the performance of the proposed scheduling algorithm, for the rtPS service class. In this experiment, the rtPS stations enter the network dynamically, one at a time, at intervals of 5 seconds, starting at time 15 seconds and a maximum of 20 SSs. We used the proportion of bandwidth 0.93:0.05:0.02 to W, T and R, respectively. The stations begin to transmit data after the network entry procedure is completed and they maintains the data transmission until the end of the simulation, in the time 160 seconds. The evaluation parameter considered is the average delay, calculated at intervals of 5 seconds.

We compared the performance of our mechanism with that of the FIFO (First-First-out) scheduler. In this scheduler, the queues are served based on time arrival of requests, unlike the proposed mechanism that prioritizes the stations whose lengths of the queues are larger. The FIFO scheduler was chosen because it is easy to implement in the framework of existing NS-2 module.

As can be seen from the graph shown in Figure 7, the proposed scheduling algorithm performs better than FIFO algorithm, when the network is saturated, i.e., the utilization of reserved bandwidth W is near 100%. This situation happens from 80 seconds onwards up to 160 seconds. This is due to the fact that in this situation the rtPS stations can not have some of their connections scheduled in current frame, which causes an increase in the queue stations, because at least one connection will be scheduled to the next frame. The reason for the superior performance of the proposed algorithm is that it prioritizes the SSs whose queue length is the longest, which does not happen in the FIFO algorithm. The main factor contributing to the increase of the delay is exactly the packet accumulation in the queues of SSs. When the network is lightly congested (15 to 75 seconds), i.e., when all connections are scheduled in current frame, no packet accumulation in the queue is observed and, hence, the performance of both algorithms is similar. The Figure 8 shows the throughput performance of the proposed scheduling algorithm in comparison with the FIFO policy. The reason for this behavior is precisely the fact that, in the time instant 80 seconds, the network is not able to schedule all the flows in the same TDD frame, which causes a substantial average delay of rtPS connections in the network, which proportionally decrease the average throughput.

2) Experiment 2: In another experiment we tried to check the influence of W variation in respect to rtPS delay and nrtPS throughput. For this, we added 8 nrtPS connections between 15 and 22 seconds (one connection by second) and every five seconds one rtPS connection is added in the network. We performed three simulations for each W value. As shown in Figure 9, the delay curve becomes more pronounced when using smaller W values. This is justified because when we use lower values for W the proposed scheduling mechanism performs the preemption faster compared to using larger values for W. Thus, when preemption occurs, i.e, from the moment that some connections have to be scheduled in the next TDD frame, the average size of the queues at the SSs increases very quickly.

As shown in Figure 10, we observe that when the value of W decreases, although this causes an increase in the delay of the rtPS connections, the throughput of nrtPS service flows increases proportionately. This is justified by the fact that, as the portion $W + T$ is constant over time, if the reservation W decrease by a factor k, i.e, became $k * W$, where $0 < k < 1$, the reservation T suffered a proportional increase of $1 - k$, i.e, $(T + W) * (1 - k)$. This means that, regardless the value assigned to W, since $T \neq 0$, there will be no bandwidth starvation in the nrtPS service flows.

B. Performance Evaluation of the Proposed CAC Mechanism

1) Experiment 3: In this experiment our goal was to verify the performance of rtPS service class with and without the proposed CAC scheme. For this purpose, we considered only rtPS stations, of which up to 20 stations enter in the network at simulation intervals of 5 seconds.

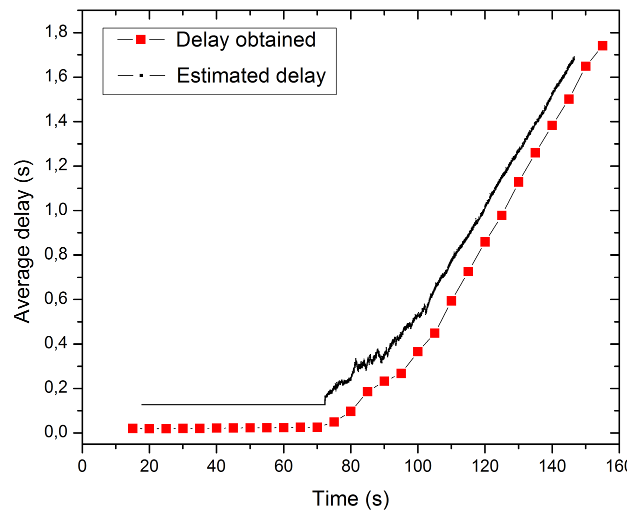

Fig. 11. Delay performance comparison between the predicted delay in the network in relation to the delay measured.

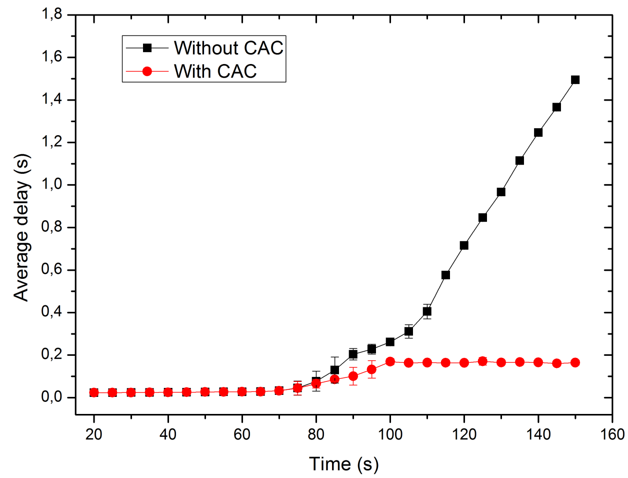

Fig. 12. Delay performance of the rtPS scheduler comparison with and without the proposed CAC scheme.

Firstly, in this experiment we evaluate the delay performance of the proposed prediction mechanism, according to equation (5). As shown in the graph in the Figure 11, the delay performance in the network is close to the predicted delay by our model for each fixed interval of 20 ms. This demonstrates that the proposed CAC mechanism is able to accurately estimate the value of network delay, which is a key factor to making the correct decision.

In other evaluation study, the rtPS delay threshold was set at 200 ms, according to Table I. The Figure 12 shows that the proposed CAC scheme limited the maximum delay to 200 ms. This CAC action prevented the addition of new connections, when the network is saturated, impair the QoS for existing connections in respect to delay parameter.

2) Experiment 4: Finally, in this experiment we evaluate the performance of the proposed CAC with respect to the CAC based on minimum rate. For this, we considered only rtPS flows added to the network after 15 seconds of simulation

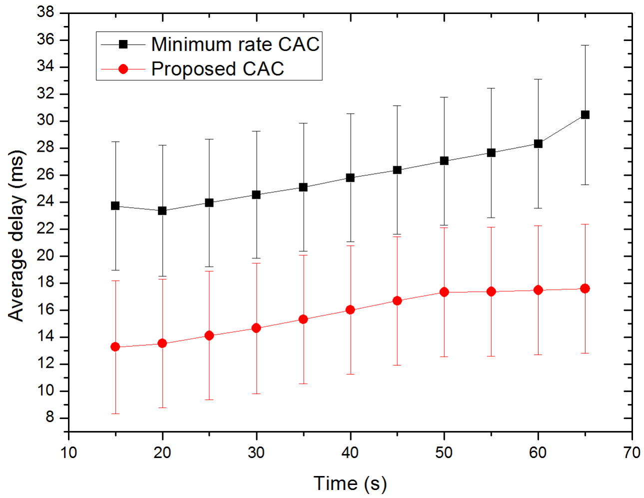

Fig. 13. Delay performance of CAC mechanism comparison with minimum rate CAC.

(one rtPS connection each five seconds). In this simulation experiment, we analyzed only the average delay parameter until such time that body mechanisms of CAC began blocking new connections, approximately at time 70 seconds. As shown in the Figure 13, although the two mechanisms have fulfilled its role in blocking the entry of new connections when the network is already saturated, the proposed CAC mechanism allows lower average delay in the network when compared with the CAC based on minimum rate. The reason for this is that the CAC based on predictive delay has greater control on the delay parameter in respect to the CAC policy based on the minimum rate, which considers only the number of connections in the network with respect to the parameters of the minimum rate. As we can see from the graph in Figure 13, when we use the CAC mechanism based on minimum rate the average delay is higher than with the proposed CAC, because it does not take into account the delay requirements in the admission process (it considers only the amount of connections in the network).

VII. Conclusions

In this paper, we presented a proposal for QoS provisioning in IEEE 802.16 standard, which consists of an uplink scheduling algorithm that works in conjunction with a bandwidth reservation mechanism and an admission control scheme based on network delay prediction. Both scheduling and admission control algorithms use the queue size information in the SSs in order to determine the priority packet and the CAC decisions, respectively. The obtained simulation results demonstrated that the proposed schemes are able to satisfy the maximum delay requirement for real-time class and improve both network throughput and delay performance for real-time and non real-time service classes. The proposed CAC mechanism was compared with a policy for admission control based on minimum rate and showed a superior performance compared to the average delay parameter.

References

[1] IEEE Standard, *Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*. IEEE Std., Rev. IEEE Std 802.16-2009, 2009.

[2] J. F. Borin (2010), *Mechanisms for quality of service provision in IEEE 802.16 networks*. Ph.D Thesis. State University of Campinas, Brazil.

[3] A. Bacciola et al, *IEEE 802.16: History, status and future trends*. Computer Communications, vol. 33, pp. 113-123, 2009.

[4] M. Castrucci et al, *Connection admissioin control in WiMAX networks*. Proceedings of the International Conference on Information and Communication Technologies, ICT'08, Paris, France, 2008.

[5] T. Tsai et al, *CAC and packet scheduling using token bucket for IEEE 802.16 networks*. Journal of Communications, vol. 1, pp. 30-37, 2006.

[6] H. Fen et al, *Performance analysis of a reservation based connection admission scheme in 802.16 networks*. Proceedings of the Global Communications Concerence, Exhibition and Industry Forum, IEEE GLOBECOM, San Francisco, United States, 2006.

[7] K. Tsang et al, *Admission control scheme for mobile WiMAX networks*. Proceedings of the International Symposium on Consumer Eletronics, ISCE'07, Dallas, United States, 2007.

[8] A. Mohamed et al, *Call admission control in wireless networks: a comprehensive survey*. IEEE Communications Surveys, vol. 7, pp. 50-68, 2005.

[9] Network Simulator 2, *http://www.isi.edu/nsnam/ns/* - NS-2 - 2010.

[10] C. Hoymann, *Analysis and performance evaluation of the OFDM-based metropolitan area network IEEE 802.16*. Computer Networks, vol. 49, pp. 341-363, June, 2005.

[11] E. Castilho et al, *An uplink scheduling and CAC algorithms for IEEE 802.16 networks*. Proceedings of the Intenational Workshop on Telecommunications, IWT'11, Rio de Janeiro, Brazil, pp. 94-100, 2011.

[12] IEEE 802.16 module for NS-2, National National Institute of Standards and Technology, *http://www.antd.nist.gov/seamlessandsecure.shtml*, 2011, last visited in april, 1.

[13] R. Prasad and F. J. Velez, *WiMAX Networks: Techno-Economic Vision and Challenges*, Springer, 2011.

Eduardo Castilho Rosa was born in Mauá - SP, Brazil, in 1986. He received his Bachelor's degree in Computer Science from Universidade Federal de Goiás (UFG) in 2007 and Master's degree in Electrical Engineering from Universidade Federal de Uberlândia (UFU) in 2011. He is currently working as professor of higher education in Instituto Federal do Triângulo Mineiro (IFTM). His research interests include performance analyses of wireless communication networks and QoS provisioning.

Paulo Roberto Guardieiro is a Full Professor at Universidade Federal de Uberlândia, Brazil, where he works since 1978. He received his B.E. degree in Electrical Engineering from Universidade Federal de Uberlândia (UFU) in 1978, the degree of M.E.E. from Instituto Tecnológico de Aeronáutica (ITA) in 1984 and the Ph.D. degree in Electrical Engineering from Universidade Estadual de Campinas (UNICAMP) in 1991. He is coordinator of the Computer Networks Laboratory of Electrical Engineering Faculty of UFU and member of the Brazilian Society of Telecommunications (SBrT). His research interests include mobile communications, multicast, QoS guarantees and DTN's.