Delay-bound Admission Control for Real-time Traffic in Fourth Generation IMT-Advanced Networks based on 802.16m

Neeraj POUDYAL[1], Ha Cheol LEE[2], Yong Jin KWON[1], Byung Seub LEE[1]
[1]*Korea Aerospace University, 100 Hangungdae Gil, Hwajeon Dong, Gyonggi Do, South Korea*
[2]*Yuhan University, 636 Gyeongin Ro, Bucheon Shi, Gyonggi Do, South Korea*
neeraj.poudyal@gmail.com

***Abstract*—In this paper a novel schedulability criteria is developed to provide Quality of Service (QoS) guarantees in terms of both minimum available bandwidth and maximum tolerated packet delay as required by the real-time traffic class. The contribution makes use of a measurement based admission control scheme at the base station of the 802.16m based 4G IMT–advanced network by considering the effects of various kinds of delays including the channel access delay, queuing delay and MAC layer transmission delay on the system's end to end delay. The paper also provides a way for the mobile station to proactively increase the chances of success of bandwidth grants by predicting in advance whether its bandwidth request will be approved by the base station, and then modifying or suspending its bandwidth request in case the chances of success is not favorable at that instant.**

***Index Terms*—IEEE 802.16m, Admission Control, Delay bound, QoS, IMT-Advanced**

I. INTRODUCTION

IM-Advanced Technology (IMT-Advanced) [1][2]) is a set of advanced requirements that will have to be present in the fourth generation (4G) networks. 802.16m [3], which is an evolution of WiMax systems based on earlier 802.16a [4] through 802.16e [5] standards, is one of the main technologies competing for the 4G networks. As such, 802.16m supports all the requirements of IMT-Advanced such as all-IP networking, seamless handovers, location-based services, high mobility and interoperability with future technologies including competing 4G systems such as LTE-Advanced [6-7]. At its core, 802.16m uses time-tested technologies such as Orthogonal Frequency Division Multiplexing (OFDM), Orthogonal Frequency Division Multiple Access (OFDMA) [8] and Multiple Input Multiple Output (MIMO) [9] to guarantee robust signal propagation with high data rates. One of the main features of these networks is to provide differentiated Quality of Service(QoS) to different application types through their six QoS classes - Unsolicited Grant Service (UGS), Real Time Polling Service (rtPS), Non Real Time Polling Service (nrtPS), Extended Real Time Polling Service (ertPS) and Best Effort (BE), and Adaptive Granting and Polling Services (aGPS) [3] ,[5]. The rtPS class, which is the focus of this paper, is used for real-time variable data rates traffic such as multimedia that has a set of QoS requirements such as minimum bandwidth, maximum tolerated latency and traffic priority. QoS provision is mainly provided through network admission control. Most works on admission control provide either bandwidth-only admission control, which means they are unable to provide delay or other QoS guarantees, or they are based on probabilistic analysis, which means these admission control are not able to provide deterministic bounds forcing a part of traffic to miss their QoS requirements. Our paper presents a novel measurement-based admission control policy to satisfy both the minimum bandwidth and maximum delay guarantees deterministically as required by some real-time applications. Additionally, in current implementations, mobile stations are not aware of the queue at the base station and stubbornly keep sending their bandwidth request (and QoS requirements) even when the base station cannot provide the service. This paper further proposes changes to the messaging system of 802.16m to help the mobile station predict in advance whether its bandwidth request will be successful before making the bandwidth request. Depending on the result of this prediction, the mobile station can make adjustments in its bandwidth and delay requirements and thus have more chance of a successful transmission in the next frame.

Guaranteeing the stringent QoS requirements for bandwidth and delay guarantees is met by scheduling policy in the network. Although providing delay guarantees is an important subject in processor scheduling area [10-17], its applicability in wireless networks has been quite limited. Unlike processors which have deterministic delay bounds, 802.16m networks are characterized by a degree of randomness in traffic and channel propagation models. In these networks the signal to noise ratio of received signal changes dynamically, which in turn switches to more or less robust modulation and coding scheme altering both the rate and delay of the network frequently. The insights gained in processor scheduling are nonetheless useful to build statistical envelope or bounds for delay as is used in some models such as [18] to achieve QoS in the network. [18] also provides a schedulability test in WiMax for QoS guarantees using different schedulers such as EDF[14] for rtPS, Fair Queuing[15],[16] for nrtPS and priority scheduling for UGS connections. Other approaches have been attempted by using queuing theory [19], statistical measurements [20], and probabilistic analysis [21-23] of delay factors. A Markov chain analysis of uplink subframe for polling is provided in [24]. Connection admission control based on number of calls is provided in [19] using both threshold-based and queue-aware approaches, but without the delay guarantees. Statistical admission control

Digital Object Identifier 10.4316/AECE.2011.01005

based on the frame occupancy is provided in [25]. To analyze end-to-end delay, it is important to analyze the delays during the contention, bandwidth request and channel access. Bandwidth request scheme has been extensively studied in [22],[25-26]. A probabilistic analysis of deadline driven admission control scheme was proposed in [17], which presented a general purpose EDF scheduling for non-preemptive flows. A comparison of performance of Region-Full and Region-Focused schemes can be found in [26]. This paper [26] also presents transmission probability, bandwidth efficiency and delay results analytically and through simulation. A latest and in-depth analysis of a bandwidth request with delay regulation has been done in [25]. The paper [25] introduces a novel approach for control knob to set the target delay and adjust the utilization accordingly. In [25], the author uses the use of step, linear and non linear functions to provide uplink scheduling with delay regulation in subscriber stations rather than the base station in a distributed manner. Our approach, on the other hand is to provide delay guarantees by using the measurements at the base station.

II. Quality Of Service In IEEE 802.16m

IEEE 802.16m IMT-Advanced system defines Advanced Base Station (ABS) and Advanced Mobile Station (AMS) as the base station and mobile station that support IEEE 802.16m. IEEE 802.16m system operating in a Time Division Duplexing (TDD) mode defines a 20ms super frame divided equally into four frames of time duration T_f =5ms each. The QoS parameter sets are pre-defined or negotiated between the ABS and the AMS during the service flow setup/change procedure. The downlink consisting of DLMAP, ULMAP and data bursts is broadcast by the ABS to all the AMSes in the system, which are received and decoded by all the AMSes. They contain the channel descriptors and synchronization information for the downlink and the uplink intervals. Similarly the uplink interval defined in the UL-Map is used by the AMSes to perform initial ranging (IR), requesting bandwidth (BR), or sending their uplink data. After an AMS requests the bandwidth, depending on the scheduling conditions and QoS availability, the ABS grants the bandwidth in full amount, partially up to the available bandwidth, or altogether rejects the request where transactional atomicity is required.

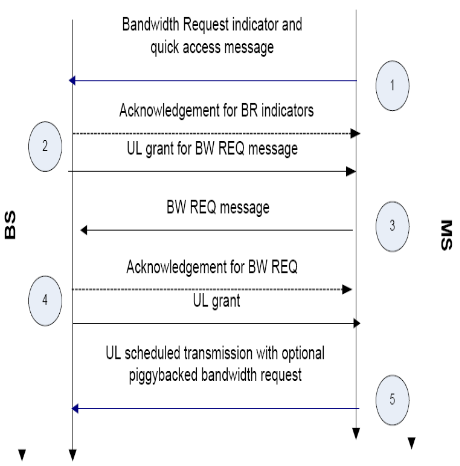


Figure 1. Bandwidth Request, Grant and Uplink [3]

A five way regular procedure or a three step quick access method may be employed for channel access (Fig.1) [3]. Bandwidth is generally granted through polling mechanism for rtPS class. The polling mode normally utilizes 3-step procedure (bandwidth request, bandwidth grant, and send data). The polling interval for rtPS class is $n*T_f$ where n=1,2… as defined by the operator. This is the maximum amount of time before an AMS has to wait before it can make a bandwidth request in its next uplink. In the following downlink, if the bandwidth in the ABS is sufficient to allow access at the negotiated maximum latency, the ABS sends a bandwidth grant via ULMAP message. If not, the bandwidth request is rejected. If there are M mobile stations, the maximum number of bandwidth requests that can be made per polling interval (n frames) is M.

III. Admission Control

1. Implementation Overview

We assume an ABS serving a number of AMSes. Although IEEE 802.16m makes a provision of Admission Control and scheduler, the details of the implementation are not defined by the standard and it is left to the vendor to decide the best solution. We fill the gap by adding our own uplink scheduler and admission control (Fig. 2). Our implementation adds some new modules to the standard 802.16m implementation, without modifying the standard.

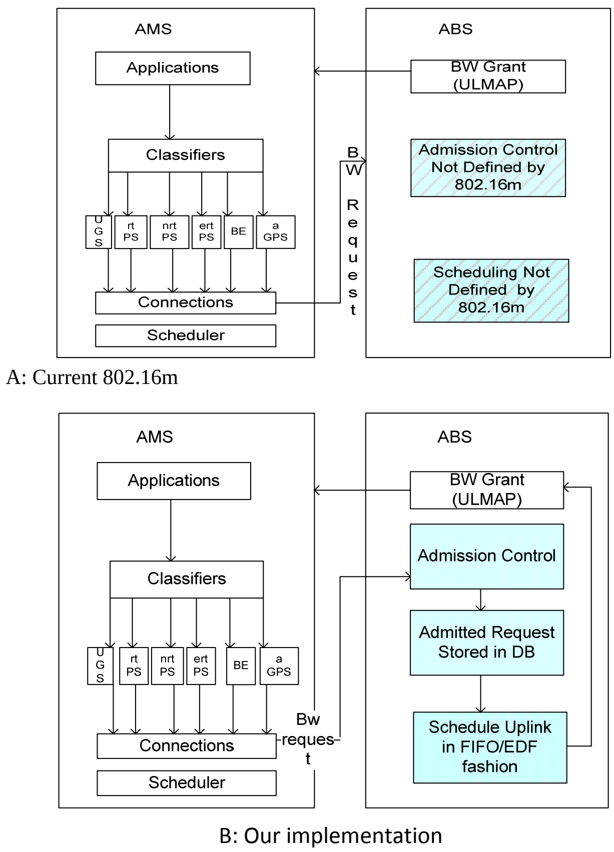


Figure 2. Major changes to 802.16m ABS (Addition of new scheduling and admission control)

Whenever a bandwidth request is made, a procedure at the base station is invoked. If the admission control criteria are met, the request entry is inserted in request table, otherwise it is rejected. The rtPS class requires guarantees in terms of minimum sustained rate and maximum tolerated delay, which is what we are going to answer in this paper. We also use some changes to messaging and AMS scheduler to optimize the channel utilization, and AMS queuing delay. For simplicity we keep the analysis at a single IEEE 802.16m system. In section IV (4), we provide an example of extending our analysis to a full multi-hop heterogeneous system with operator-level constraints such as maximum user rate, minimum sustained rate, as well as any number of additive 'unforeseen' constraints.

2. End to End Latency Analyses

End to end delay is a combination of polling delay, channel acquisition delay, queuing delays at the AMS and the ABS, and MAC-layer transmission delay. We ignore packet preparation and packet processing delay in our analysis assuming them to be negligible and not within the focus area of the network and MAC layers.

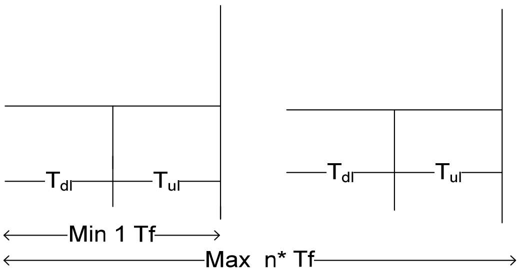


Figure 3. Polling Delay

Fig.(3) shows the maximum time the AMS with a non-empty rtPS queue has to wait before it can send its bandwidth request. This delay depends on the polling interval defined by the implementer or the operator of IEEE 802.16m system in use, and it is in the multiple of frame duration, $n*T_f$ *(n=1,2..)*. For this analysis we define polling delay *Dp* equal to the polling interval.

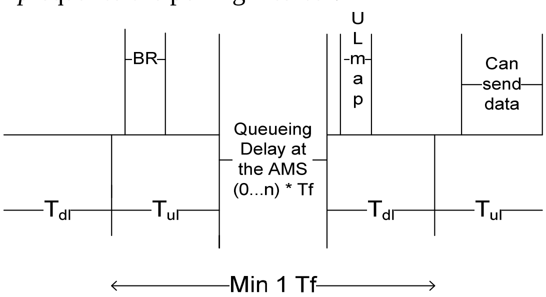


Figure 4. Channel Acquisition Delay

Fig.(4) shows the delay occurred during the acquisition of channel. When the AMS is polled, it sends a bandwidth request during its uplink, T_{ul} (and the ABS receives it at the same uplink, T_{ul}). If there is enough bandwidth in the next uplink, the ABS grants the channel its immediate downlink; otherwise the ABS has to wait for *k* number of frames before it can grant the channel to this connection. The minimum delay for channel acquisition (*Dc*) is one T_f. If the bandwidth is rejected by the ABS, the AMS has to retry again in the next poll. For the sake of simplicity we define channel acquisition delay as the minimum time T_f required to send the bandwidth request and acquire the channel by the AMS. The period the AMS has to wait before actually receiving the ULMAP depends on the capacity of the IEEE 802.16m system, and is treated separately.

Fig.(5) shows the delay that occurs at the MAC layer after source AMS is allowed to send its data. Upon receiving the ULMAP in the downlink, the source AMS sends its data during the subsequent uplink (and the BS receives it at the same time T_{ul}).

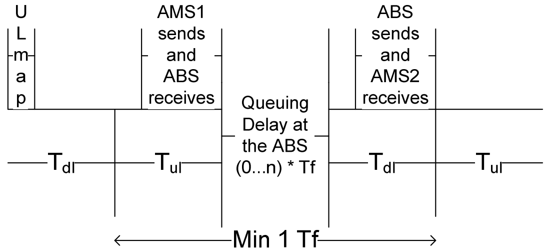


Figure 5. MAC delay, after the source AMS sends

As discussed before, the rate at which this transmission takes place depends on the robustness of the modulation scheme, defined by the AMC. Note that most previous papers [12-13],[15],[17],[19] do not consider AMC into account whereas our algorithm measures the data rate considering the effect of AMC implicitly. If there is a bottleneck in the downlink, meaning there is already data to send from previous transmissions because of the smaller downlink to uplink ratio, the data received by the ABS must wait *k* number of frames before it can be sent to the destination AMS. If on the other hand the downlink is free, the source ABS can immediately send the data in its next downlink frame, T_{dl}. From fig. 2, we find at the minimum, the MAC layer delay in delivering (D_m) requires $1\ T_f$ time frame. The end to end delay must be in the multiple of frame duration (T_f) as all the delays occurring during channel acquisition or transfer, are at least of T_f duration.

Observation 1. Provided there is no bottleneck in the downlink, the minimum end to end delay D_{min} for a successful request is given by:

$$D_{\min} = D_P + D_C + D_M = (n+2)T_f \qquad (1)$$

Observation 2. Now we can consider the effects of Queuing delay *Dqss* in the calculation of End to end delay *D*. Let *Dqss* be in the '*m*' multiple of frames such that $Dqss=m*T_f$. Then *D* is given by:

$$D = D_P + D_C + D_{qss} = (n+m+1)T_f \qquad (2)$$

Queuing delay is the time required to serve the bandwidth request, *TBr*, excluding channel access time, given by

$$T_{Br} = Br / R \qquad (3)$$

where *Br* is the requested bandwidth and *R* is the actual (measured) rate of the 802.16m system at the modulation rate of the channel and m is simply the time *TBr* converted into multiple of frames,

$$m = Ceil\left(T_{Br} / T_f\right) \times T_f \qquad (4)$$

(3) considers the traffic from only one AMS. The schedulability condition in section 3.4 considers the effect of traffic from all other AMSes.

3. Rate of IEEE 802.16m

TABLE 1: Adaptive Modulation and Coding Scheme example

SNR	Modulation	Coding	Useful Bits Per Symbol (ubps)
6.0	QPSK	½	192*2*1/2=192
8.5	QPSK	¾	192*2*3/4=288
11.5	16 QAM	½	192*4*1/2=384

15.0	16 QAM	¾	192*4*3/4=576
19.0	64 QAM	2/3	192*6*2/3=768
21.0	64 QAM	¾	192*6*3/4=864

The rate of transmission between the ABS and AMS is chosen by the Adaptive Modulation and Coding Scheme depending on the received Signal to Noise Ratio (SNR) of the AMS. Depending on the Modulation and coding, the useful bits per symbol (ubps) is calculated as shown in example table 1. The time to transmit each OFDM symbol with sampling frequency *fs* and cyclic prefix ratio *G* can be used to derive the rate of the system.

$$Symbol\ Time = (1/[fs/NFFT]) \times (1+G)$$

$$Rate = Number\ of\ symbols \times ubps \qquad (5)$$

$$Rate = ubps/((1/[fs/NFFT]) \times (1+G))$$

(5) provides a theoretical system rate. Since *ubps* depends on the state of each AMS, they have to be summed up to get the system rate. In our implementation, we use measured rate rather than theoretical rate for accuracy.

4. Schedulability Tests

Define AND operator Φ such that a Φ b returns TRUE if conditions a and b are both satisfied, else returns FALSE.

Schedulability Criteria 1. A new rtPS connection with a new bandwidth request Br_i, and delay request D_i , is schedulable with delay guarantees in a 802.16m system with C already admitted bandwidth requests if

$$Br_i \le B_{tol} - (\sum_{k=1}^{C} Br_k - \Upsilon_k) - \sum_{j=1, j\ne k}^{N} \Upsilon_j$$

$$\Phi$$

$$D_i \ge (n+1)T_f \ +$$

$$Ceil\left(1/T_f \ (\sum_{k=1}^{C} Br_k/R_k + Br_i/R_i) \right) \times T_f$$

$$D_i \in N \times T_f, N = 3, 4, \ldots \qquad (6)$$

Terminology

Br= bandwidth request. Br is removed after it is served

Br_k=Bandwidth request of already admitted connection k

Br_i=Bandwidth request of the new connection i

B_{tol}=Total system bandwidth capacity (measured)

Υ_k = Minimum sustained rate of station k

N = number of active AMS

Di= Maximum Delay requested by the new request i

C= total number of pending requests, each request is removed from C when it is served

R_k = the rate of the total 802.16m system measured with the modulation rate of the station *k*. For example if the measured rate(capacity) of the IEEE 802.16m system is 10 Mbps at 64 QAM(3/4) and 8 Mbps at 64 QAM(1/2), and the modulation rate of station k is 64 QAM(1/2) as dictated by its SNR, then R_k = 8Mbps.

Proof: The first inequality in (6) (first operand of the operator Φ) is to check for utilization bound, whereas the second inequality ensures that the delay bounds are not exceeded. The proof of the first inequality is trivial. To prove the second inequality, we will explain the meaning of each term. The first term is the sum of polling and channel acquisition delay. The second term is the time in frames it takes to serve all the previous backlog requests, and the current request. It is the queuing delay at the AMS including the MAC layer transmission delay. The maximum bandwidth that can be transferred by an AMS is equal to the transmission time (delay) multiplied by the transmission rate. The term is divided by Tf, “ceil”ed, and again multiplied by Tf. This is done just to convert the queuing delay into the multiple of frames. As an example let’s say we have 10 frame duration delays, at 5ms frame duration and a 10 Mbps channel modulated rate. The maximum bandwidth that can be served at these conditions is 10*.005*10 =0.5Mb. In other words, in 10Tf we can serve 0.5 Mb of data.

IV. Implementation Details And Results

1. Changes to ABS

The two inequalities (operands of Φ) of (6) can directly be used to construct an admission control policy. A continuous form of the (6) is easier to implement as the admission control. Two such schedulability testing techniques are presented in the following section for continuous bandwidth updating - the first one (7) uses a timer and tracks the time, and the second one (8) tracks the bandwidth served. We use the second technique (8) in our implementation.

Schedulability Criteria 2 (Continuous Form). A new rtPS connection with a new bandwidth request Br_i, and delay request D_i , is schedulable with delay guarantees in a 802.16m system with C already admitted bandwidth requests if

$$Br_i \le B_{tol} - (\sum_{k=1}^{C} Brem_k - \Upsilon_k) - \sum_{j=1, j\ne k}^{N} \Upsilon_j$$

$$\Phi$$

$$D_i \ge (n+1)T_f \ +$$

$$Ceil\left(1/T_f \ (\sum_{k=1}^{C} Brem_k/R_k + Br_i/R_i) \right) \times T_f$$

$$- \ T_{elapse}$$

$$D_i \in N \times T_f, N = 3, 4, \ldots \qquad (7)$$

where T_{elapse} is the time elapsed since the first request Br_f $(f \in C)$ at the head of the rtPS queue. This second inequality of (7), when compared to (6), requires more timer overhead but frees up the resources continuously as the traffic are being served. Equivalently if we continually keep track of the remaining bandwidth, (bandwidth requested minus served), which we do in our implementation, the inequality can be changed to:

$$Br_i \le B_{tol} - (\sum_{k=1}^{C} Brem_k - \Upsilon_k) - \sum_{j=1, j\ne k}^{N} \Upsilon_j$$

$$\Phi$$

$$D_i \ge (n+1)T_f \ +$$

$$Ceil\left(1/T_f \ (\sum_{k=1}^{C} BRem_k/R_k + Br_i/R_i) \right) \times T_f$$

$$D_i \in N \times T_f, N = 3, 4, \ldots \qquad (8)$$

where $BRem_K$ is the pending bandwidth of the k^{th} request remaining to be served. In other words it is the original bandwidth that was requested by the request k minus the bandwidth served by now. The conceptual implementation of scheduling and admission control policies is shown in Fig. 6 (actual implementation is done using a linked list instead of database for storage).

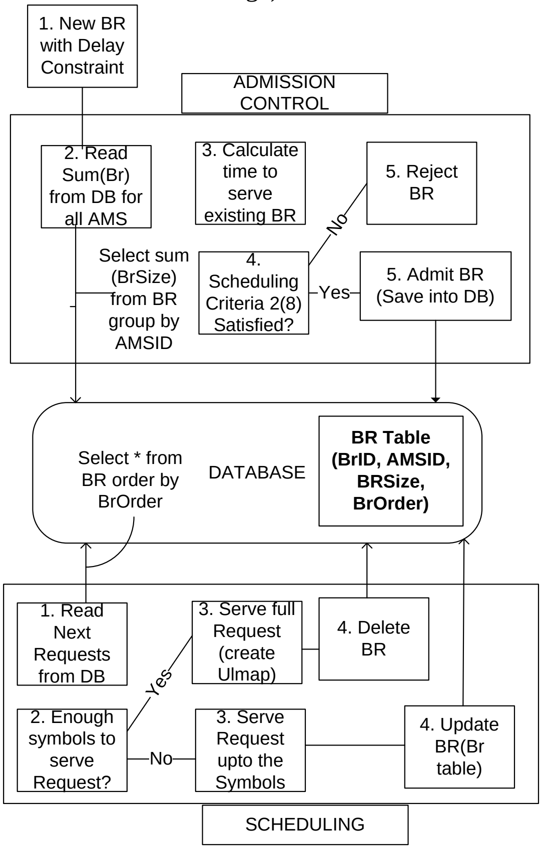


Figure 6. Conceptual View of Implementation Details (AC and Scheduling)

Bandwidth Request table (BR table) stores all the admitted requests along with the IDs of the AMSes the requests are associated with. When a new request arrives, the admission control retrieves the sum of rtPS bandwidth request size for each AMS from the BR table. Next, delay is calculated by dividing this amount with the current modulation rate of each AMS. Admission control is then applied using (8), and if the new bandwidth request is accepted it is saved into the BR table in the database. Scheduling is independent of admission control policy. Scheduling picks the next request from the BR table, sees if the symbols are enough to serve the request in full amount, and if so, the scheduler serves the request then removes the request entry from the BR table. If the symbols are only sufficient to serve partial request, the scheduler grants bandwidth (ULMAP) for the request equal to the amount possible to be served by remaining symbols, then updates the remaining bandwidth in the database. This way the remaining bandwidth is always updated for each AMS every frame.

Simulation result (Fig. 7a) shows that for different delays during full utilization, our admission control misses almost no deadlines. Simulating the same scenario in the 802.16m system without the admission control, we see the percentage of deadlines missed by the traffic is inversely proportional on the duration of the deadline itself. This also means that the bandwidth served by our admission control has to trade off with the delay (Fig. 7b). For the traffic with lower delay bounds, the system can serve less bandwidth. The unused bandwidth can be used by the traffic having a higher delay requirement. The effect on the queuing delay on the received packets is shown in Fig. 8. The delay of AC packets shows the standard deviation of only 7 resulting in a more regular delay curve whereas the standard deviation of non AC packets is 22 resulting in a non-regular (less predictable) delay curve.

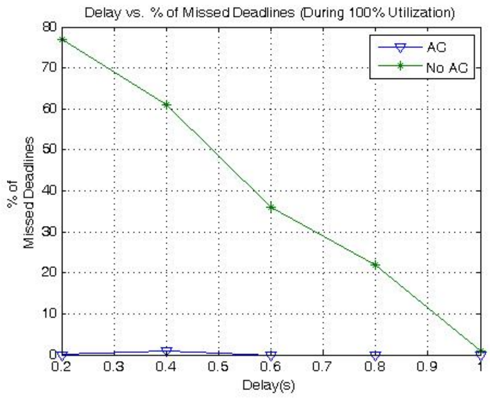


Figure 7a. Average % of Deadlines Missed (full utilization) for different delays

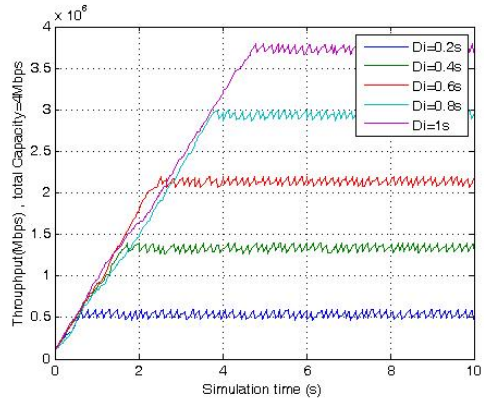


Figure 7b. Throughput-Delay tradeoff

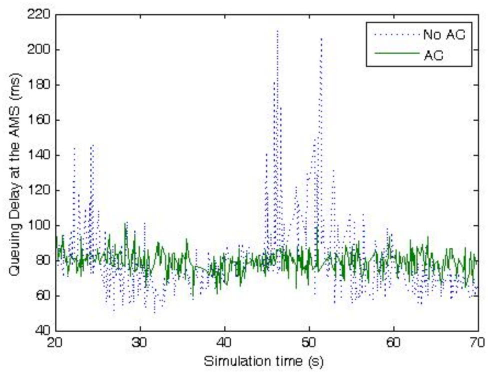
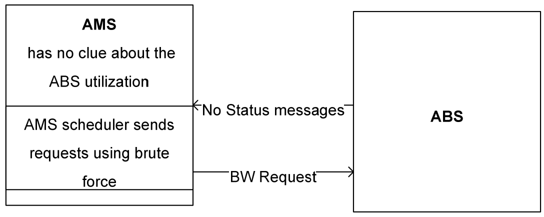


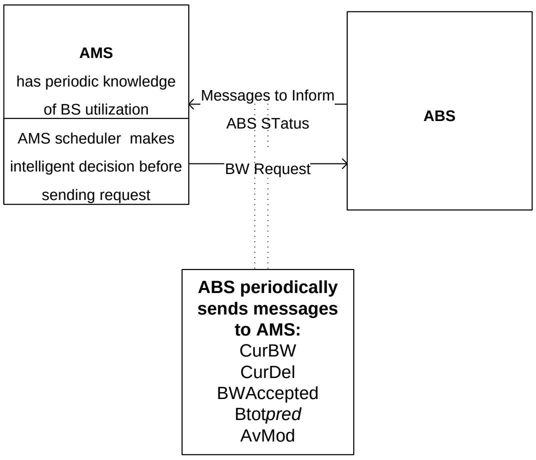
Figure 8. Queuing Delay of the received packets

2. Changes to Messaging of IEEE 802.16m



a. 802.16m AMS unaware of the utilization status at the ABS, makes brute-force bandwidth requests

The AMS of current 802.16m has no way of knowing the system utilization status at the base station (Fig. 9a). The result is that it keeps asking for bandwidth during every n frame(s) even when the base station is not able to admit it. The size of the queue can grow indefinitely, and the applications have no way of predicting how much bandwidth or delay the base station will allow the AMS to send. If the AMS is periodically informed about the utilization and delay status of the ABS, the AMS can inform its applications and adjust their rate and delay requirements so that it has higher chances of admitting its traffic (Fig. 9b).



b: Our AC aware AMS knows utilization status at the BS and makes intelligent bandwidth requests

Figure 9. Addition of messages to 802.16m to allow AMS to make intelligent bandwidth requests

To provide the AMSes with some clue about the admission control at the base station, the base station can broadcast the calculated *CurBW*, *CurDel* such that

$$CurBW = Btol - \left(\sum_{k=1}^{C} Brem_k - \Upsilon_k\right) - \sum_{j=1,\, j\neq k}^{N} \Upsilon_j \qquad (9)$$

$$CurDel = T_f \times Ceil\left(\frac{1}{T_f}\sum_{k=1}^{C} Brem_k / R_k\right) \qquad (10)$$

at every i (i=1,2..) frame in its downlink. The ABS also needs to inform the AMS whether its last bandwidth request was accepted. In addition, two values, the total amount of bandwidth request granted in the current polling interval $Btot_{pred}$, and average modulation rate Av_{mod} can be provided by the ABS.

The first application of such information is in the prediction of delay. If an AMS expects its bandwidth request Br_m to be admitted into the network then the best, worst and average case maximum delay D_m that AMS *m* can request successfully is given by [Br_m/R_m], [$(B_{tot}-Br_m)/R_{av} + Br_m/R_m$] and [$((B_{tot}-Br_m)/R_{av} + 2\,Br_m/R_m)/2$] respectively expressed in the multiple of frame duration (In other words, $D_m \in n*T_f$). The term R_m is the rate of IEEE 802.16m using modulation rate of AMS *m*, R_{av} is the average rate of IEEE 802.16m system, and B_{tot} is the measured capacity of IEEE 802.16m system.

3. Changes to AMS Scheduler of 802.16

Let $B_{pred(m)}$ be the predicted bandwidth that other AMSes will be granted before granting bandwidth to station *m* during the current polling interval.

$$B_{pred}(m) \cong \sum_{q=1}^{m-1} Br_q \qquad (11)$$

Note that Bpred only considers the new bandwidth that is requested by the other AMSes during this polling interval, not considering the bandwidth that is already granted until the last polling interval (which is CurBW). Let Dpred be the predicted new transmission time (delay) consumed by the new requests and calculated as:

$$D_{pred}(m) \geq T_f \times Ceil\left(B_{pred}(m)/R_{av} \times 1/T_f\right)$$
$$\cong \sum_{q=1}^{m-1} D_q \qquad (12)$$

Before sending its bandwidth request, the AMS m with bandwidth request Brm (with delay request Dm) can approximate whether the request will be accepted by using (13).

$$Br_m \leq CurBW - B_{pred}(m)$$
$$\Phi$$
$$D_m \geq CurDel + (n+1)T_f +$$
$$Ceil\left(1/T_f \times Br_m/R_m\right) \times T_f + D_{pred}(m)$$
$$Assuming$$
$$D_{pred} \geq T_f \times Ceil\left(B_{pred}(m)/R_{av} \times 1/T_f\right),$$
$$where\ R_{av} = Average\ Modulation\ Rate$$
$$D_m \in N * T_f,\ N = 3, 4 \ldots,$$
$$B_{pred}(m) \cong \sum_{q=1}^{m-1} Br_q\,,$$
$$D_{pred}(m) \cong \sum_{q=1}^{m-1} D_q \qquad (13)$$

To precisely predict Bpred is an impossible task. Therefore only an approximation can be made by the AMS scheduler by observing the previous trends and setting Bpred, value equal to the last Btotpred sent by the ABS.

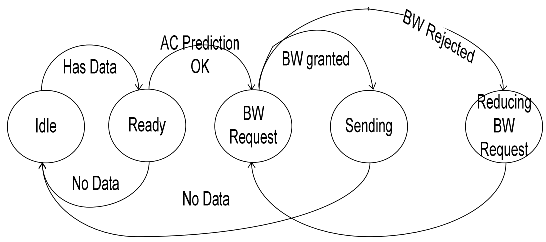


Figure 10: State transition diagram of AMS BW request

If the last request was not successful, B_{pred} can be decreased by a factor f in the next requests until it is finally accepted. It can then be increased with the same factor, much like a TCP adaptive windowing system [29]. The state diagram of SS for this purpose is shown in Fig 10. The estimation errors E_β and E_δ are the differences between the predicted and actual values of *Bpred* and *Dpred* respectively:

$$E_\beta = B_{pred}(m) - \sum_{q=1}^{m-1} Br_q\,,$$
$$E_\delta = \sum_{q=1}^{m-1} D_q - D_{pred}(m) \qquad (14)$$

Negative values of E_β and E_δ are undesirable and may mean the transmission won't be successful. Large positive values result in the underutilization of the system. The use of (13) is most effective when the base station is in full utilization.

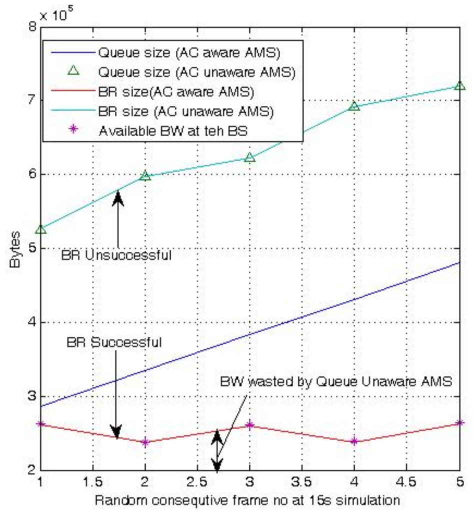


Fig 11: Bandwidth Request and AMS Queue size

Simulation result (Fig. 11) shows that adding our scheme to the AMS can save the system bandwidth of IEEE 802.16m, as well as significantly shorten the size of the AMS queue. We randomly pick certain frames after around 15 seconds of simulation with arrival rate λ=1/0.0001 packets/second, packet size of 500 bytes, and reserved rtPS bandwidth of 5 Mbps. We show five random consecutive frames as frame 1, 2 etc in the X-axis in fig. 11. The AMS that is unaware of the base station AC status sends the same amount of bandwidth (BR) as its queue size. However, as shown in fig. 11, the ABS may not have sufficient bandwidth to serve that bandwidth and it will be unsuccessful. The bandwidth request of an AC-aware AMS is generally close to the remaining bandwidth, and thus the request is usually accepted if there are no secondary delay constraints that can't be met. The difference in the queue sizes of the two schemes shows that AC-aware scheme is more effective in the sense that it wastes much less bandwidth. Note that in the case of nrtPS or BE traffic [3],[4],[5], the delay-bound admission control is unnecessary and the ABS, instead of rejecting the request, automatically grants the bandwidth equal to its available bandwidth .

3.1. Successful Bandwidth Request

(13) is an approximation by AMS m that its bandwidth request will be successful. True probability of success is however given by.

$$
\begin{aligned}
&= \Pr(Br_m \le CurBW - B_{pred}(m)) \\
&\times \Pr(D_m \ge CurDel + (n+1)Tf \\
&+ Ceil\left(\tfrac{1}{Tf} \times \tfrac{Br_m}{R_m}\right) \times Tf + D_{pred}(m)) \\
&\times \Pr(B_{pred}(m) \le \sum_{q=1}^{m-1} Br_q) \\
&\times \Pr(D_{pred}(m) \ge \sum_{q=1}^{m-1} D_q)
\end{aligned}
\qquad (15)
$$

3.2. Queue Stability at the AMS

To ensure the queue length of AMS does not increase indefinitely, the AMS has to ensure its arrival rate λ is less than or equal to its departure rate μ .

$$
\Pr\left(\text{Queue won't grow indefinitely}\right)
$$

$$
= \Pr\left(\int_{t=0}^{\infty} \mu dt - \int_{t=0}^{\infty} \lambda dt \ge 0\right) \qquad (16)
$$

where λ and μ are arrival and departure rates respectively. Thus the condition for the expected stability of SS queue over F number of frames is given by:

$$
\begin{aligned}
&\sum_{f=1}^{F} \mu_f = [\Pr(Brm_f \textit{ is admitted})] \times Brm_f \\
&= \sum_{f=1}^{F} [\Pr((Brm_f \le CurBW - B_{predf}(m) \\
&\qquad \Phi \\
&Dm_f \ge CurDel + (n+1)T_f + \\
&Ceil\left(\tfrac{1}{T_f} \times \tfrac{Br_m}{R_m}\right) \times T_f + D_{predf}(m)) \\
&\qquad = TRUE)] \times Brm_f \\
&\ge \sum_{f=1}^{F} \lambda_f
\end{aligned}
\qquad (17)
$$

where $\{Br_{mf}, D_{mf}\}$ are the requested bandwidth and delay by AMS m during frame f, and $\{B_{predf(m)}, D_{pref(m)}\}$ are the predicted bandwidth and delay requested during frame *f* by all other AMSes whose requests are admitted before admitting request by m, and λ_f and μ_f are arrival and departure rates during the frame *f*. (18) can be used by the AMS to inform its traffic generators to speed up or slow down depending on the queue size.

4. Multi-hop, Multi-constraint, and User Rate Control Extension

We can easily extend our admission control criteria for a multi-hop network with multiple additive QoS constraints and operator parameters for rate control. Let user station *k*'s minimum sustained rate be Υ_k and maximum user rate be Γ_k . Furthermore we have j^{th} additive constraint weights *aij(p)* in each hop *p* of path *P* , and a user defined j^{th} constraint criteria *Aij(P)* such that for all *j*, *Aij(P)* is at least the sum of *aij(p)*. The criterion for such schedulability is:

$$
Br_i \le MIN\left(Btol - \left(\sum_{k=1}^{C} BRem_k - \Upsilon_k\right) - \sum_{\substack{j=1,\\ j\ne k}}^{N} \Upsilon_j\right.
$$

$$
\left., MIN(Bp \mid p \varepsilon P) - \left(\sum_{k=1}^{C} BRem_k - \Upsilon_k\right) - \sum_{\substack{j=1,\\ j\ne k}}^{N} \Upsilon_j\right)
$$

$$
\Phi
$$

$$
D_i \ge (n+1)T_f + \sum_{p=1}^{P} \frac{Br_i + \sum_{k=1}^{C} B\,\mathrm{Re}\,m_k}{Bp}
$$

$$+ Ceil\left(\frac{1}{T_f} \left(\sum_{k=1}^{C} \frac{BRem_k}{R_k} + \frac{Bri}{\Gamma i} \right) \right) T_f$$

$$\Phi_{j=1}^{MAX\ CONSTRAINTS} \quad A_{ij}(P) \geq \sum a_{ij}(p)$$

$$Di \in N \times T_f, N = 3, 4, ...$$

$$Bp = bandwidth\ of\ a\ link\ p\ in\ P \tag{18}$$

Note that delay is simply one case of A. Other examples of additive constraint Aij may be number of hops and packet loss probability.

I. CONCLUSION

In this paper, we developed schedulability criteria for admitting real time traffic into IEEE 802.16m based fourth generation IMT- Advanced Network, to ensure QoS guarantee in terms of both minimum bandwidth and maximum tolerated delay. Each request for bandwidth is checked by the admission control policy that either approves or rejects the packet depending on whether the system bandwidth and delay constraints are sufficient to meet those requirements. This admission control policy is able to meet almost 100% of the deadlines, which is verified using simulations. The AMS and the messaging system are modified so that the AMS is aware of the queue at the ABS which increases the utilization of the ABS while simultaneously decreasing the queue size of the AMS. Implementation details regarding the changes to IEEE 802.16m AMS, ABS and messaging system are provided and results are verified. We also provide a full multi-hop, multi-constrained extension with rtPS user-level capacity parameters.

REFERENCES

[1] "Requirements related to technical performance for IMT-Advanced radio interface(s)", ITU, 2008,

[2] "Guidelines for evaluation of radio interface technologies for IMT-Advanced", ITU, 2008,

[3] "IEEE 802.16m System Description Document (SDD) modification 2009-09-24", Website, IEEE, 2009

[4] "IEEE 802.16-2004 IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems", IEEE, 2004

[5] "IEEE 802.16e-2005 IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems. Amendment 2 for physical and MAC layers for combined fixed and mobile operations in licenced bands", IEEE, 2005

[6] "LTE-Advanced", 3GPP, 2009

[7] I. Papapanagiotou , D. Toumpakaris , J. Lee and M. Devetsikiotis, "A Survey on Next Generation Mobile WiMAX Networks: Objectives, Features and Technical Challenges", IEEE Communications Surveys and Tutorials, Volume 11, No. 4, IEEE, 2009

[8] H. Yin, S. Alamouti, "OFDMA: A Broadband Access Technology", Sarnoff Symposium, IEEE, 2006

[9] Q. Li, X. E. Lin, J. Zhang, W. Roh, "Advancement of MIMO technology in WiMax: from IEEE 802.16d/e/j to 802.16m", IEEE Communication Magazine, Vol. 47, Issue 6, IEEE, 2009

[10] C. L. Liu and J. W. Layland, "Scheduling Algorithms for Multiprogramming in Hard Real-time Environment", Journal of the Association for Computing Machinary, Vol. 20, No. 1, January 1973, pp. 46-61

[11] Q. Zheng, K.G. Shin, "On the Ability of Establishing Real-Time Channels in Point-to-Point Packet Switched Networks", IEEE transactions on communications, Volume 42, Issue 234,IEEE, 1994, pp. 1096-1105.

[12] J. Liebeherr, D. E. Wrege and D. Ferrari, "Exact Admission Control for Network with a Bounded Delay Service", IEEE/ACM Transactions on Networking, Vol. 4, No. 6 , IEEE, December 1996, pp 885-901

[13] S. Baruah, and T. Baker, "Schedulability Analysis of Global EDF", Real-Time Systems, Volume 38 , Issue 3, Kluwer Academic Publishers, Norwell, MA, USA, 2008, pp. 223-235

[14] V. Sivaraman, F. Chiussi., "Providing End-to-End Statistical Delay Guarantees with Earliest Deadline First Scheduling and Per-Hop Traffic Shaping", Proc. of IEEE INFOCOM, 2000

[15] S. Lu, V. Bharghavan, and R. Srikant.,"Fair scheduling in wireless packet networks", IEEE/ACM Transactions on Networking", vol. 7. No 4.IEEE, 1999, pp. 473-489

[16] M. Shreedhar, G. Varghese, "Efficient Fair Queuing Using Deficit Round Robin", IEEE/ACM transactions on networking, Vol 4, No 3, IEEE, 1996, pp. 375-385

[17] R. Simon and T. Znati, "A Probabilistic Analysis of Admission Control Policies for Deadline Driven Service Disciplines", Computer Networks: The International Journal of Computer and Telecommunications Networking, Volume 48 , Issue 2, pp 137-154 , June 2005

[18] K. Wongthavarawat and A. Ganz, "Packet Scheduling for QoS support in IEEE 802.16 broadband wireless access systems", International Journal of Communication Systems, Vol. 16, pp. 81-96,Feb 2003

[19] D. Niyato and E. Hossain, "Connection Admission Control Algorithm for OFDM Wireless Networks, IEEE GLOBECOM, MO, USA, 28 Nov-2 Dec 2005

[20] J. Y. Lee, K. B. Kim, "Statistical Connection Admission Control for Mobile Wimax Systems", Wireless Communications and Networking Conference, 2008

[21] R. Iyengar, K. Kar and B. Sikdar, "Scheduling Algorithms for Point-to-Multipoint Operation in IEEE 802.16 Networks", Proceedings of 2nd Workshop on Resource Allocation in Wireless Networks (RAWNET), April 2006.

[22] R. Iyengar, V. Sharma, K. Kar and B. Sikdar, "Analysis of Contention based Multi-Channel Wireless MAC for Point to multipoint Networks", Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, Buffalo, June 2006

[23] R. Iyengar, P. Iyer and B. Sikdar, "Delay Analysis of 802.16 based Last Mile Wireless Networks", Proceedings of IEEE Globecom, 2005

[24] B.J. Chang, C. M. Chou and Y. H. Liang, "Markov Chain Analysis of Uplink subframe in polling-based WIMAX networks", Computer Communications, Vol. 31, Issue 10, pp. 2381-2390, June 2008

[25] E. C. Park, "Efficient Uplink Bandwidth Request with Delay Regulation for Real-time Service in Mobile WiMax Networks", IEEE Transactions on Mobile Computing, Vol.8, No. 9, IEEE, September 2009, pp 1235-1239

[26] J He, Z. Tang and H. Chen, "Performance Comparision of OFDM Bandwidth Request Schemes in Fixed IEEE 802.16 Networks", IEEE Communication Letters, Vol. 12, No. 4, IEEE, April 2008, pp 283-285

[27] "NIST WiMax Module, Seamless and Secure Mobility", National Institute of Standards and Technology.

[28] A. Belghith and L. Nuaymi, "Design and Implementation of a QoS-included WiMAX Module for NS-2 Simulator", First International Conference on Simulation Tools and Techniques for Communications, Networks and Systems, SimuTools 2008, Marseille, France, 3-7 March 2008.

[29] A. Belghith, L. Nuaymi, and Patrick Maillé, "Pricing of differentiated-QoS services WiMAX networks", IEEE Global Communication Conference, IEEE GLOBECOM 2008, New Orleans, LA, USA, 30 November – 4 December, pp. 1-6

[30] A. Allman, V.Paxson, W.Stevens, "TCP Congestion Control", RFC 2581, http://tools.ietf.org/html/rfc2581

[31] A. K. Parekh and R. G. Gallager. "A generalized processor sharing approach to flow control in integrated services networks: The single node case". IEEE/ACM Trans. on Networking, 1(3):344-357, 1993.