Enhanced Unsolicited Grant Service (eUGS) for WIMAX Networks

Ahmad Ammouri & Jamil N. Ayoub
EE Dept. College of Engineering, University of Jordan, Amman, JORDAN
ayoub@ju.edu.jo

Abstract: - This paper investigates the WiMAX standard (IEEE 802.16) and focuses on the QoS architecture of its network MAC layer. The different kinds of available scheduling services suitable for VoIP are studied, and a new scheduling service called Enhanced Unsolicited Grant service (eUGS) is proposed. Performance of (eUGS) was studied by simulation, and the results show a QoS improvement over the other scheduling services that carry VoIP.

Key-Words: - MAC layer, scheduling service, UGS, Simulation.

1Introduction

Worldwide Interoperability of Microwave Access, WiMAX, is a wireless data network which can provide Internet service designed to cover wide geographical areas serving large numbers of users at low cost. The IEEE 802.16 standard defines WiMAX as the standard being adopted worldwide by manufacturers to insure inter-operability of equipment. A WiMAX network has very advanced features like fast air link, symmetric downlink/uplink capacity, and a flexible resource allocation mechanism. These features enable WiMAX to meet QoS requirements for a wide range of data services and applications.

In the WiMAX Medium Access Control (MAC) layer, QoS is provided via service flows. Service flow is a unidirectional flow of packets that is provided with a particular set of QoS parameters. Before providing a certain type of data service, the Base station (BS) and Subscriber station (SS), first establish a unidirectional logical link between their peer MACs called a connection. The MAC peer then associates packets traversing the MAC interface into a service flow to be delivered over the connection. The QoS parameters associated with the service flow define the transmission ordering and scheduling on the air interface [1].

MAC scheduler determines the data handling mechanisms for data transport on a connection. Each connection is associated with a single scheduling service. Each data service is associated with a set of QoS parameters that quantify aspects of its behavior.

The MAC scheduler supports four scheduling services: Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-real-time Polling Service (nrtPS), and Best Effort (BE). Each of these supported scheduling services include some mandatory QoS parameters when the scheduling service is enabled for a service flow.

The UGS could be used to support real-time data streams consisting of fixed-size data packets issued at periodic intervals, such as T1/E1 and VoIP without silence suppression. The mandatory QoS service flow parameters for this scheduling service are Maximum Sustained Traffic Rate, Maximum Latency, Tolerated Jitter, and Request/Transmission Policy. The Minimum Reserved Traffic Rate parameter, if present, will have the same value as the Maximum Sustained Traffic Rate parameter.

The rtPS is designed to support real-time data streams consisting of variable-sized data packets that are issued at periodic intervals, such as moving pictures experts group (MPEG) video. The mandatory QoS service flow parameters for this scheduling service are Minimum Reserved Traffic Rate, Maximum Sustained Traffic Rate, Maximum Latency, and Request/Transmission Policy.

The remaining nrtPS and BE scheduling services arenot suitable for real time data as voice, and will not be considered in this paper.

Concerning QoS for voice traffic over the MAC layer, each one of the above types of services has different mechanisms to transfer their traffic. As a result, voice over these different types of services will have different quality under different conditions.

IEEE 802.16 standard provide extensive bandwidth allocation and QoS mechanisms. Details of scheduling and reservation management are left unstandardized, which provide an important mechanism for vendors to differentiate their equipment, and open up research efforts to address these issues. [2].

The scheme proposed in this paper intends to improve voice traffic over the MAC layer of the WiMAX. It will study the different kinds of scheduling services that are available to carry VoIP traffic and their mechanisms to maintain the QoS of the transmitted voice. As a result of this study, amendments, denoted as eUGS, to some of the scheduling services will be suggested to handle VoIP traffic. The validity of this eUGS is studied by simulation.
There is a large body of work done by different researchers to improve scheduling and management procedures on the MAC layer of the WiMAX. In [3], a new scheduling algorithm was suggested to deal with voice activity on the uplink of IEEE 802.16d/e to increase the number of voice users that could be served. Adaptive contention resolution for VoIP services in the IEEE 802.16 networks was studied in [4], and presented analytical calculations for the adaptive parameters to ensure VoIP QoS while controlling contention resolution.

An Integrated QoS control architecture for broadband wireless access systems was studied in [5], and a new QoS architecture was suggested to enhance a mapping rule and a fast signaling mechanism for providing both InterServ and DiffServ over 802.16 network.

2 Enhanced Unsolicited Grant Service

Voice traffic that will be considered in this paper is voice that is processed by modern encoders and prepared to be carried over IP networks, normally denoted as (VoIP). Voice traffic can accept up to 1% loss without effecting it's Quality [6]. Due to this nature, all kinds of VoIP traffic are sent over User Datagram Protocol (UDP). However, voice quality is sensitive to delay and jitter.

Depending on the Vocoder used, Voice traffic is generated with fixed packet length [7]. Since a voice stream is not continuous, a normal call uses about 35.13% of the channel time in a full duplex system [8].

2.1 Voice Over MAC Layer

According to the 802.16 standard, voice could be carried over UGS or rtPS [2]. UGS service is designed to allocate continuous fixed-size bandwidth (BW) to the SS on a periodic basis. An SS will not use request/transmission mechanism, and delay will be fixed since the BW is reserved from the beginning of the connection. Also traffic will not suffer from  jitter since delays are controlled.

As voice traffic is not continuous, there will be silent periods through a call duration. During silent periods the BS will allocate BW for the SS while the SS will not be using it. Such allocated, or reserved, BW will be lost and neither the SS nor other SSs will use it. This constitutes about 64.87% of the allocated BW for any voice call [8]. This loss is the main disadvantage of carrying voice over UGS. This loss will decrease the available resources for the network. And with high loads, the network will become congested, thereby increasing the blocking rate. Such waste of network resources will lower the QoS for voice as well as for other traffic type services.

In voice traffic transmitted over rtPS type of service, an SS will send a request to the BS to allocate BW for it. This means that during silence periods, the SS will not ask for such allocation, and no BW will be lost, except that which is allocated to poll the SS.

These BW requests are considered as overheads (lost BW), since there is no traffic carried over it. Since the size of the lost BW due to BW requests is smaller than the size of the lost BW in the case of UGS, rtPS mechanisms will be more Bandwidth efficient than UGS.

However, in rtPS type of service, the allocation of bandwidth is not guaranteed, and a BS will allocate BW to an SS only when BW is available. This means that jitter and delay, at peak hours, will be very high.

2.2 Proposed (eUGS)

The main concept of this eUGS scheduling service is to divide the SSs into silent and non silent SSs, and then allocate BW grants to all non silent SSs according to normal UGS operation, and allocate a number of equal Floating Grants (FG) to the silent SSs to be used by any of them that decides to become non silent. Operation of eUGS can be summarized as follows:

1- Connection initiation is the same as in UGS. An SS requests fixed size bandwidth with a specific rate to meet its need. In case of voice traffic, the size of this request will be consistent with the coding type in the (Vocoder). The BS will allocate this bandwidth and announce it in the Up-link map and the Down-link map.

2- The Down-link map is built as in UGS, and according to the MAC layer in the WiMAX network. However, the Up-link map will have the same structure except for some modifications. This map is built by announced BW allocations for non silent SSs, and an announced number of equal FGs for silent SSs.

3- Before becoming silent, an SS sends “I am silent” message, causing the BS to stop allocating any BW grants to it.

4- Switching to non silent mode, after a silence period, an SS will use any available FG to send its traffic. The BS then checks its connection ID, and resumes allocating BW grants for it in the next Up-link maps according to the size of the bandwidth that was reserved at the initiation of the connection.

Any specifications needed for eUGS, will obviously be related to the FG. For one thing, the size of the FG should equal the maximum bandwidth allocated to any one of the eUGS connections, so that an SS will not fragment its packets to fit on the FG. The other issue is related to the possibility of collision among SSs desiring to use the FG at the same time.
2.3 Collision Resolution In eUGS

In this section, we present a technique to deal with the possibility of collision when using an FG. In this technique when there is a collision on the FG, the BS will allocate extra floating grants in the next Up-link map according to following relation:

\[ N_{\text{Floating}} = 2^{n+1}, \quad N_{\text{Floating}} \leq N_{\text{Silent}} \quad \cdots (1) \]

Where “N_{\text{Floating}}” is the number of FGs, “N_{\text{Silent}}” is the number of silent SSs, and n is the number of continuous collisions (collisions that followed each other on the next FG). Once there is no collision on the FG, the counter of continuous collisions will be reset, and the number of FGs will go back to be one.

When an SS finds that there is more than one FG, the SS will randomly select any one to send its packet. Upon collision, an SS will drop its current packet and send the next packet also using a randomly selected FG.

An alternative method would be for the BS to divide the SSs into subclasses and allocate one FG to each subclass.

3 Simulation of eUGS

To test the performance of the eUGS scheduling service, and to validate its advantages over the other current services, a simulation study over the MAC layer of the WiMAX network was designed and run under different network loads. The services simulated included UGS, rtPS, UGS and UGS/AD as in [3].

The structure of the simulation algorithm for eUGS is shown in Fig. 1. During the Initiation stage, the simulator is configured with some parameters to make the simulator run on conditions very close to those of the real network under different loads. These parameters include: duration of simulation, average number of users, time between two up-link maps, mean call duration, and generation time between two groups of users.

In the Generate Users step, each generation time, the system generates a number of users N, equal to a Poisson random number with mean equal to the average number of users [9]. Moreover, to simulate a practical network, where new users do not all come at the same time, the simulation creates users randomly at different times within the duration of the generation time. For each user, an actual call duration T, is generated equal to an exponential random variable with average mean equal to the mean call duration. Also, a born time recording when an SS starts a call is generated for each generated user.

Generated traffic for each user is assumed normal VoIP, in accordance with G729 Codec given in Table 1 [8]. The model for on and off periods is an exponential random variable with parameter for active periods equal to T_{on} and parameter for off periods equal to T_{off}.

<table>
<thead>
<tr>
<th>Codec</th>
<th>Packet Size (Bytes)</th>
<th>Packet Interarrival (ms)</th>
<th>T_{on} (Sec)</th>
<th>T_{off} (Sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G729</td>
<td>70</td>
<td>30</td>
<td>0.352</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Table 1: VoIP Traffic

For the Build Uplink Map, the BS will check all registered users and allocate BW grants for them, and allocate an FG to the silent users. The Up-link map shape may look as in Table 2.

| User | User | User | User | Floating | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>36</td>
<td>61</td>
<td>43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Up-link Map

The Serve SSs step is illustrated in the SS Up-link Transmission Algorithm given in Fig.2. The step when Collision is detected, is resolved as explained above invoking Eq. (1) as necessary.

4 Simulation Results

The results to be displayed are obtained from running a simulation of the MAC layer of a WiMAX network carrying VoIP traffic over many scheduling services including: UGS. UGS/AD, rtPS, and eUGS. The values for the initiation parameters are assumed as follows: Simulation run time = 120 minutes, Generation
time = 3 minutes, Average call duration = 3 minutes, and number of iterations = 240000.

4.1 Bandwidth Occupancy

Bandwidth occupancy is expressed in terms of the average number of grants utilized. And a grant represents a time slot on the up-link that carries one voice packet every 30 ms with 70 byte payload and 6 bits of overhead. Results of the simulation for BW occupancy is given in Table 3, and Fig. 3.

As evident in Fig.3, the best service schedule is eUGS, under normal loads. Increasing the load in the network will cause high collision on the FG BW, so the BS will allocate more FGs each time, and this will increase the average number of used grants and makes the UGS/AD better than eUGS when the average number of users exceed 60.

The performance of eUGS can be improved more under high network load, by dividing the number of users to subclasses, each subclass will share evenly one floating grant. Fig.4 shows the eUGS grants occupancy under different subclass sizes. From Fig. 4, the best performance is when the subclass size is between (20-30) users.

### Table 3

<table>
<thead>
<tr>
<th>Average Number of Users</th>
<th>Optimal Needed Grants</th>
<th>Average needed grants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UGS</td>
<td>rtPS</td>
</tr>
<tr>
<td>10</td>
<td>3.7973</td>
<td>10.7984</td>
</tr>
<tr>
<td>20</td>
<td>6.6327</td>
<td>18.7397</td>
</tr>
<tr>
<td>30</td>
<td>9.7829</td>
<td>27.7311</td>
</tr>
<tr>
<td>40</td>
<td>12.9546</td>
<td>36.5980</td>
</tr>
<tr>
<td>50</td>
<td>18.1418</td>
<td>51.4550</td>
</tr>
<tr>
<td>60</td>
<td>23.8334</td>
<td>67.7239</td>
</tr>
<tr>
<td>70</td>
<td>26.6057</td>
<td>75.2398</td>
</tr>
<tr>
<td>80</td>
<td>29.3820</td>
<td>83.2117</td>
</tr>
<tr>
<td>90</td>
<td>33.0686</td>
<td>93.7077</td>
</tr>
<tr>
<td>100</td>
<td>33.8665</td>
<td>95.5717</td>
</tr>
<tr>
<td>110</td>
<td>37.3460</td>
<td>105.5496</td>
</tr>
<tr>
<td>120</td>
<td>43.8574</td>
<td>124.1899</td>
</tr>
</tbody>
</table>

Fig. 3 : BW occupancy

Fig. 4 : BW occupancy for eUGS for different Subclass sizes.

![Fig. 2 : SS Up-link Transmission Algorithm](image-url)
4.2 BW Occupancy Analysis

According to voice call statistics, on average, when a user makes a voice call, he uses the circuit for about 35.13% of the call duration [8]. From Table 3, the ratio of every entry in the second column to the corresponding entry in the third column is approximately (0.3513), which justifies the above claim. Thus, we have:

\[ BW_{Optimal} = 0.3513 \times BW_{UGS} \quad \cdots (2) \]

Where \( BW_{Optimal} \) denotes the optimal number of grants needed by an ideal system, and \( BW_{UGS} \) is the number of grants needed when using UGS.

To derive a relationship between BW grants needed by rtPS and UGS, we note first that the normal time slot size of a MAC packet is 76 bytes, while the bandwidth request size is 22 bytes. We can write:

\[ BW_{rtPS} = BW_{Optimal} + \frac{22}{76} \times BW_{UGS} \quad \cdots (3) \]

\[ P_{rtPS/UGS} = \frac{BW_{Optimal} + \frac{22}{76} \times BW_{UGS}}{BW_{UGS}} \]

\[ = 0.5313 + 22/67 = 0.6408 \quad \cdots (4) \]

Where \( BW_{rtPS} \) represent grants needed by a system using rtPS, and \( P_{rtPS/UGS} \) is the ratio of needed rtPS grants to those needed by UGS. The ratio in Eq.(4) is justified by the simulation results given in Table 3, since dividing every entry in column 4 by its corresponding entry in column 3 gives approximately 0.64.

The relation between BW grant needs of UGS/AD and UGS can be derived by first noting that grant needs for UGS/AD (\( BW_{UGS/AD} \)) is equal to grant needs for non Silent SSs (\( BW_{Optimal} \)) and grant needs for "bandwidth requests" for all silent SSs. With 64.87% of all SSs silent, and with the grant needs for "bandwidth request" is equal to 22/76 of the grant needs for non silent SS, we can write:

\[ BW_{UGS/AD} = BW_{Optimal} + .6487 \times \frac{22}{76} \times BW_{UGS} \quad \cdots (5) \]

and,

\[ BW_{Optimal} + .6487 \times \frac{22}{76} \times BW_{UGS} \]

\[ P_{(UGS/AD)/UGS} = \frac{BW_{Optimal} + .6487 \times \frac{22}{76} \times BW_{UGS}}{BW_{UGS}} \]

\[ = 0.3513 + 0.6478 \times 22/76 \]

\[ = 0.5388 \quad \cdots (6) \]

Again, this result is approximately as obtained by simulation when every entry in column 5 is divided by its corresponding entry of column 3 in Table 3.

Finally, the relation between eUGS and UGS BW occupancy can be expressed in terms of a network load function, f(N) as:

\[ BW_{eUGS} = BW_{Optimal} + f(N) \quad \cdots (7) \]

This follows from the fact that \( BW_{eUGS} \) is equal to grants needed for non silent SSs, \( BW_{Optimal} \) plus grants needed for FGs. The average number of FGs depends on the collisions which are directly related to the network load, N. Hence we have Eq. (7). The values for f(N), as obtained in the simulation, are shown in Table 4 which shows its dependence on the network load.

<table>
<thead>
<tr>
<th>No. of users</th>
<th>BW optimal</th>
<th>( BW_{eUGS} )</th>
<th>f(N)</th>
<th>Collision percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3.7973</td>
<td>5.3237</td>
<td>1.5264</td>
<td>4.7400</td>
</tr>
<tr>
<td>20</td>
<td>6.6327</td>
<td>8.9125</td>
<td>2.2798</td>
<td>13.6638</td>
</tr>
<tr>
<td>40</td>
<td>12.9546</td>
<td>18.1043</td>
<td>5.1497</td>
<td>37.2163</td>
</tr>
<tr>
<td>50</td>
<td>18.1418</td>
<td>27.0681</td>
<td>8.9263</td>
<td>51.7350</td>
</tr>
<tr>
<td>60</td>
<td>23.8334</td>
<td>37.4691</td>
<td>13.6357</td>
<td>62.3625</td>
</tr>
<tr>
<td>70</td>
<td>26.6057</td>
<td>42.3206</td>
<td>15.7149</td>
<td>67.8663</td>
</tr>
<tr>
<td>80</td>
<td>29.3820</td>
<td>47.1323</td>
<td>17.7503</td>
<td>70.6175</td>
</tr>
<tr>
<td>90</td>
<td>33.0686</td>
<td>54.3119</td>
<td>21.2432</td>
<td>74.2013</td>
</tr>
<tr>
<td>100</td>
<td>33.8665</td>
<td>55.9206</td>
<td>22.0541</td>
<td>74.8688</td>
</tr>
<tr>
<td>110</td>
<td>37.3460</td>
<td>63.1781</td>
<td>25.8321</td>
<td>77.2763</td>
</tr>
<tr>
<td>120</td>
<td>43.8574</td>
<td>74.6089</td>
<td>30.7515</td>
<td>79.9588</td>
</tr>
</tbody>
</table>

4.3 BW limited system

In an actual system, where resources are limited, there is a maximum size of BW available at the BS. If the SSs' BW needs exceed the BS's available BW, then some SSs will not be allocated BW. As voice traffic, though very sensitive to delay and jitter, could accept some packets to be dropped, the SS simulation algorithm suggests dropping any packets of any non silent SS if it has no BW grant allocated for it.

When the simulation was modified to run under BW limitation, the results for BW occupancy and packet dropping rate are given in Fig.5, and Fig. 6 respectively. These figures show the relative advantages of eUGS service over the other services without adversely affecting voice QoS.
5. Conclusions

In this paper, the QoS architecture of the MAC layer of the WiMAX standard (IEEE 802.16) has been investigated. The various kinds of available scheduling services suitable for VoIP are studied, and a new scheduling service called Enhanced Unsolicited Grant service (eUGS) is proposed and its procedures explained. Performance of (eUGS) was studied by simulation, and the results show a QoS improvement over the other standard scheduling services. Also, its performance was compared with UGS/AD, as proposed in [3].

Improvements provided by eUGS include BW occupancy savings at the same level of QoS as in UGS, for both bandwidth limited or unlimited systems. Freeing more BW means an increase in network capacity to carry real time and/or non real time traffic. Also, eUGS has the option of maintaining it performance edge as the load increases, by dividing users into subclasses and allocate FGs to each subclass.

References


