Optimal Packetization Interval for VoIP Applications Over IEEE 802.16 Networks

Shehan Perera
Department of Electrical & Computer Engineering
University of Canterbury
New Zealand
shehan@elec.canterbury.ac.nz

Harsha Sirisena
Department of Electrical & Computer Engineering
University of Canterbury
New Zealand
sirisehr@elec.canterbury.ac.nz

Krzysztof Pawlikowski
Department of Computer Science
University of Canterbury
New Zealand
krys@canterbury.ac.nz

Abstract

An analysis of the impact of the packetization interval for constant bit rate traffic has been done in the context of IEEE 802.16 MAC layer. Bandwidth used for overheads which include lower layer headers as well as retransmissions at the MAC layer are considered. An optimal packetization interval selection method for delay sensitive applications such as VoIP is proposed. Enhancements to the Unsolicited Grant Service retransmission strategy are proposed to further improve delay and minimize packet loss while making efficient use of the limited bandwidth resource.

I. Introduction

THE need for mobile as well as fixed point wireless connectivity has increased greatly in recent years and as such new protocols and access technologies are emerging to satisfy this need. One of most spoken of as of late is IEEE 802.16 also known as WiMAX. This standard was originally intended for point-to-multipoint fixed broadband wireless access. It aims to provide high data rates for a large number of users in a wide area.

One of the applications which WiMAX specifically caters for is VoIP. According to the standard [1] four classes of scheduling have been defined of which unsolicited grant service (UGS) and real-time polling service (rtPS) can potentially be used to schedule uplink VoIP traffic. Even so there are several issues when it comes to this kind of service flow.

Synchronizing the grant of uplink bandwidth in either of the above schedule types can be a problem. For UGS several solutions to this problem are proposed in [2]. These solutions propose adjusting the grant time in steps ranging from 15 frames down to the maximum tolerated jitter. Smaller steps of adjustment introduce lower jitter but can take up to hundreds of milliseconds to reach the desired level of latency.

During silent periods of a voice conversation bandwidth is wasted in UGS scheduling. This wastage is minimized by using voice activity detection/silence detection at the subscriber station (SS) and informing the base station (BS) of the voice state transitions [3, 4]. This allows more voice users to be accommodated in the system than would be using "pure" UGS. Usage of rtPS scheduling has also been investigated but does have the drawbacks of added delay due to the request-grant mechanism and overhead due to unused request slots.

Since the protocol data units (PDUs) can vary considerably in size, from tens of bytes to tens of kilobytes it is important to investigate the effect on efficiency and link utilization. In [5] the optimal PDU size has been calculated for different residual bit error rates (BER) in order to minimize system overhead. This wastage of bandwidth due to overhead is more as the payload size decreases. In the case of VoIP packets with packet sizes in the tens of bytes, it is very important to consider the Orthogonal Frequency Division Multiplexing (OFDM) symbol parameters when deciding the optimal packetization interval. The packet size is directly proportional to the packetization interval.

The rest of this paper is organized as follows. Section II gives a brief overview of IEEE 802.16 and the UGS schedule type. Next we will provide an analysis of the effect of packetization interval selection and propose a method to dynamically select the best value during run time.

II. Overview of IEEE 802.16

A. IEEE 802.16 MAC Protocol

IEEE 802.16 assumes a point-to-multipoint architecture with a central base station (BS) which acts as gateway to connect the subscriber stations (SS) in the cell to other public networks. The MAC operation is based on

MAP messages transmitted periodically (once per frame) by the BS. The MAP defines the times in the downlink (DL) and uplink (UL) which are used for ranging, contention based bandwidth requests, allocated polled type bandwidth requests, DL PDUs and UL PDUs for SS to send data to the BS. Ranging is a process which is done by the SS at initial entry into the system and periodically at the request of the BS or the SS itself to optimize signal quality.

The responsibility of scheduling UL/DL data is entirely up to the BS. Depending on the Quality of Service (QoS) requirements of a particular flow it will be classified into one of the four schedule classes [6]. The MAC defines Dynamic Service Addition/Change/Deletion messages (DSA, DSC, and DSD) which are used to agree upon the flow parameters using a request/response/acknowledge (REQ, RSP, and ACK) 3 way handshake process. Out of the two classes which can be used for real time flows we will consider only UGS in this work although the results can be applicable to the other classes.

B. Operation of UGS

UGS is designed to provide fixed size data grants at periodic intervals to real time constant bit rate (CBR) like traffic flows. This reserves a guaranteed bandwidth for flows without the overhead and latency of the request grant mechanism. Since the data grants are provided on a periodic basis, the BS can estimate the application's requirement with respect to its QoS level during connection initialization. The standard defines four QoS parameters for UGS flows.

1) Maximum sustained traffic rate: defines the peak data rate, which in the case of UGS is also the minimum reserved rate.

2) Maximum latency: the delay between receiving a packet from the network layer and forwarding the packet to the physical layer at the transmitter. This is basically the time taken for the MAC layer to process the packet and get it onto the air interface.

3) Tolerated jitter: an upper bound on the amount of delay variation that can be tolerated at the application level.

4) Request/Transmission policy: defines the rules of uplink bandwidth request and PDU formatting. All forms of uplink requests are prohibited for the UGS connection.

III. Analysis of Packetization Interval

A. Relevant Equations

Consider a VoIP application which produces a voice data stream of r bits-per-second (bps). The overhead due to headers $OH_{headers}$ is the sum of the RTP, UDP, IP and MAC layers headers in bits. n_{pkt} is the packet size as seen at the MAC layer. t_{pkt} is the packetization interval of the VoIP application.

$$n_{pkt} = r \times t_{pkt} + OH_{headers} \tag{1}$$

n_{pdu} and n_{bps} are the size of the PDU and bits per symbol respectively. The ceil function rounds upwards towards the closest integer.

$$n_{pdu} = ceil\left(\frac{n_{pkt}}{n_{bps}}\right) \times n_{bps} \tag{2}$$

The symbol error rate, SER is given in equation 3. Here m gives the maximum number of bit errors which can be tolerated. PER is the packet error rate and n_{spp} is symbols per packet.

$$SER = 1 - \sum_{j=0}^{m} \binom{n_{bps}}{j} BER^{j} \left(1 - BER\right)^{n_{bps} - j} \tag{3}$$

$$PER = 1 - \left(1 - SER\right)^{n_{spp}} \tag{4}$$

Overhead due to retransmissions (in bits) is given in (5), where n is the MAC layer retransmit limit for this particular traffic class. Bandwidth used for overhead is found by dividing the total overhead by the packetization interval, (6). E_f is the efficiency of the system. *payload* being the actual voice data from the application layer.

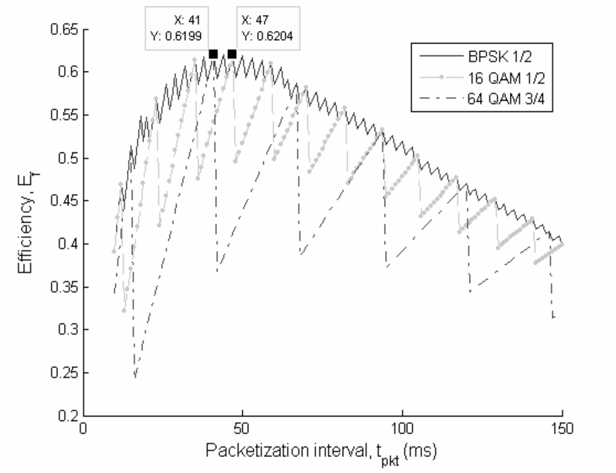
$$OH_{ret} = n_{pdu} \left\{ \sum_{i=1}^{n} \left\{ iPER^{i} \left(1 - PER\right) \right\} + nPER^{n} \right\} \tag{5}$$

$$OH_{bw} = \frac{(n_{pdu} + OH_{ret} - payload)}{t_{pkt}} \tag{6}$$

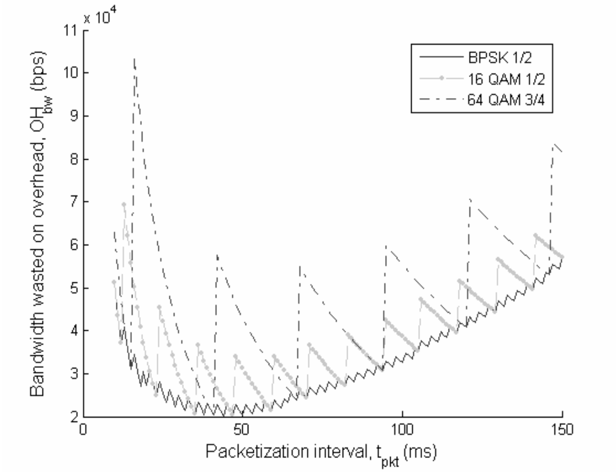
$$E_f = \frac{payload}{n_{pdu} + OH_{ret}} \tag{7}$$

When the retransmit limit n has been exceeded the packet is considered lost. This probability is P_{loss}.

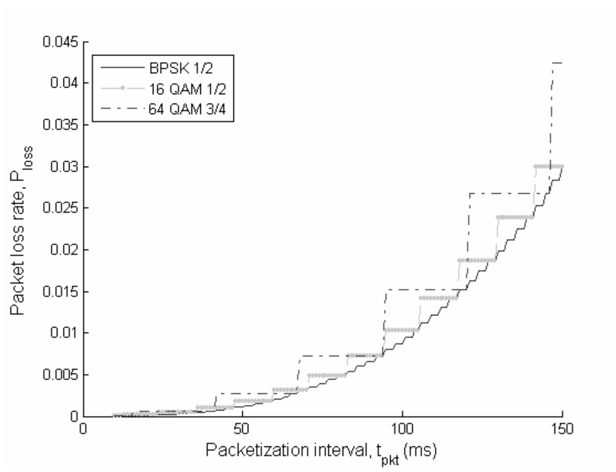
$$P_{loss} = PER^{n+1} \tag{8}$$



(a)

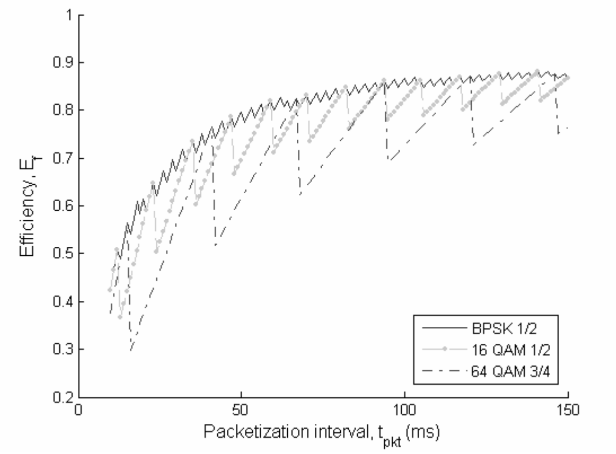


(b)

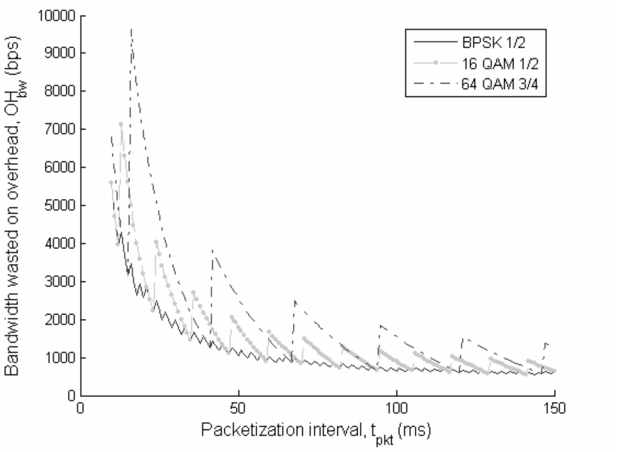


(c)

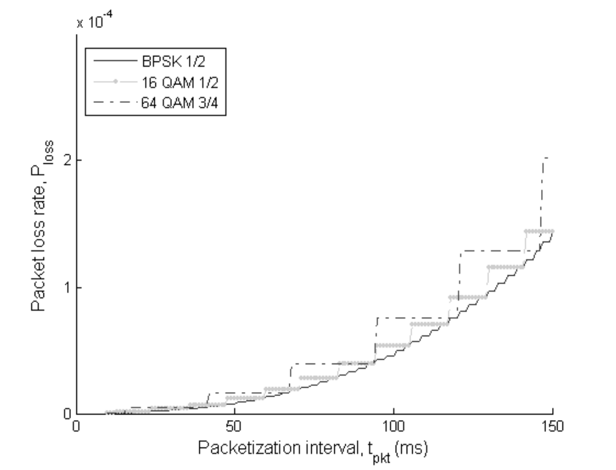
Fig. 1. (a) Efficiency of the system (b) Overhead bandwidth usage (c) Packet loss rate of a VoIP application with a BER of 1e-4



(a)



(b)



(c)

Fig. 2. (a) Efficiency of the system (b) Overhead bandwidth usage (c) Packet loss rate of a VoIP application with a BER of 1e-6

B. Sample Scenario

A sample scenario using common values for 802.16 will be considered to demonstrate the effects of t_{pkt}. n_{bps} for different burst profiles are given in Table I. We are considering the UL phase of the flow of a 256 sub carrier OFDM system. The burst profile used depends on the signal-to-noise ratio which depends on the distance from the BS. The VoIP application is assumed to produce a stream at a rate of 32kbps. The retransmission limit is set at 2 and the maximum tolerated bit errors is 0.

Fig. 1 gives the Efficiency, overhead bandwidth and packet loss rate for packetization intervals ranging from 10ms to 150ms with a BER of 10-4. The saw tooth effect is due to the transmission units being integer multiples of OFDM symbols. The most efficient intervals are shown in Fig. 1(a). For 64 QAM 3/4 this interval 41ms produces a packet loss rate of 0.4%. The next best interval would be between 11ms~15ms which has a packet loss rate of 0.05%.

When the BER is lower as in Fig. 2 the optimal packet size is larger, which is intuitive. In the context of VoIP it is not possible to select the largest possible packetization interval satisfying the QoS packet loss limit. We also need to stay within the latency bounds of the flow. By selecting an interval between 38ms and 40ms, an efficiency of about 70% can be achieved for the 64 QAM case. With an interval between 42ms and 47ms the same efficiency can be achieved for the 16 QAM case.

It is also clear from both Fig. 1(b) and Fig. 2(b) that a difference of a few milliseconds can increase the overhead bandwidth up to 10s of kbps which can be a few times the bandwidth of the voice application.

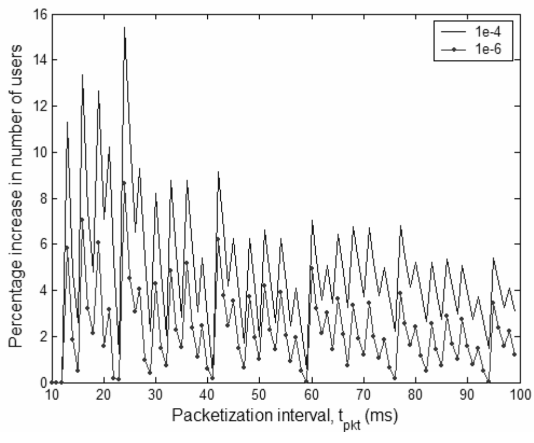


Fig. 3. Gives the percentage increase in the number of users for a fixed amount of UL resources. Two values of BER are compared.

As given in [5] based on the SNR requirements of the different modulation schemes (or burst profiles) the cell area can be classified into annulus regions. The boundaries mark the change to a lower modulation.

TABLE I

Modulation Scheme	Bits per OFDM symbol (n_{bps})	Percentage of total area (a_i)
BPSK 1/2	96	39.40
QPSK 1/2	192	20.56
QPSK 3/4	288	27.95
16 QAM 1/2	384	4.10
16 QAM 3/4	576	5.15
64 QAM 1/2	768	0.92
64 QAM 3/4	864	1.92

Different modulation schemes used and the bits per OFDM for each of them.

The area of these annuluses as a percentage of the total cell area is given by a_i, where for example a_3 represents the region which can use any one of the three lowest modulation schemes. The 3rd column of Table I contains these values. We assume that the total number of SSs in the cell is uniformly spread out. $b_{r,i}$ is the effective bit rate of a SS in the ith annulus for a randomly chosen t_{pkt}. $b_{o,i}$ is the effective bit rate of a SS in the ith annulus for an optimally chosen t_{pkt} (which is lower than the randomly selected t_{pkt}). n_r and n_o are the number of users in the system for random t_{pkt} and optimal t_{pkt}, using the same amount of resources measured in OFDM symbols. The ratio of n_o:n_r given in (9) is plotted as a percentage increase in Fig. 3.

$$\sum_{i=1}^{7} n_r a_i \left({b_{r,i}}/{n_{bps,i}} \right) = \sum_{i=1}^{7} n_o a_i \left({b_{o,i}}/{n_{bps,i}} \right)$$
$$n_o / n_r = \left. \sum_{i=1}^{7} a_i \left({b_{r,i}}/{n_{bps,i}} \right) \middle/ \sum_{i=1}^{7} a_i \left({b_{o,i}}/{n_{bps,i}} \right) \right. \qquad (9)$$

The gain due to optimal selection of t_{pkt} is more at higher BERs due to the increased retransmission overhead. A bigger gain can be achieved by using an optimal t_{pkt} which is larger than the random t_{pkt} but this will increase the latency. The average latency l_{avg} is given by (10). This includes an additional component t_{pkt} in the summation which accounts for the lag due to packetization.

$$l_{avg} = \sum_{i=0}^{n} \left[\left(t_{pkt} + \frac{t_{pkt}}{2} + iT_f \right) (1 - PER) PER^i \right] \qquad (10)$$

T_f is the frame duration. For low BER values (<1e-4) latency is roughly equal to $1.5t_{pkt}$. Without latency minimizing enhancements we have considered the MAC service delay to be half of the packetization interval. At every talk spurt the starting point could randomly fall anywhere in the range (0, t_{pkt}] from the current grant position.

IV. Implementation Scheme

At the start of the service flow the initiating SS will send a DSA_REQ message to the BS. If the receiver is also a part of an IEEE 802.16 cell it too should follow the same procedure. To do this the application layer of the SS must communicate with the MAC layer to alert of the beginning of the voice stream.

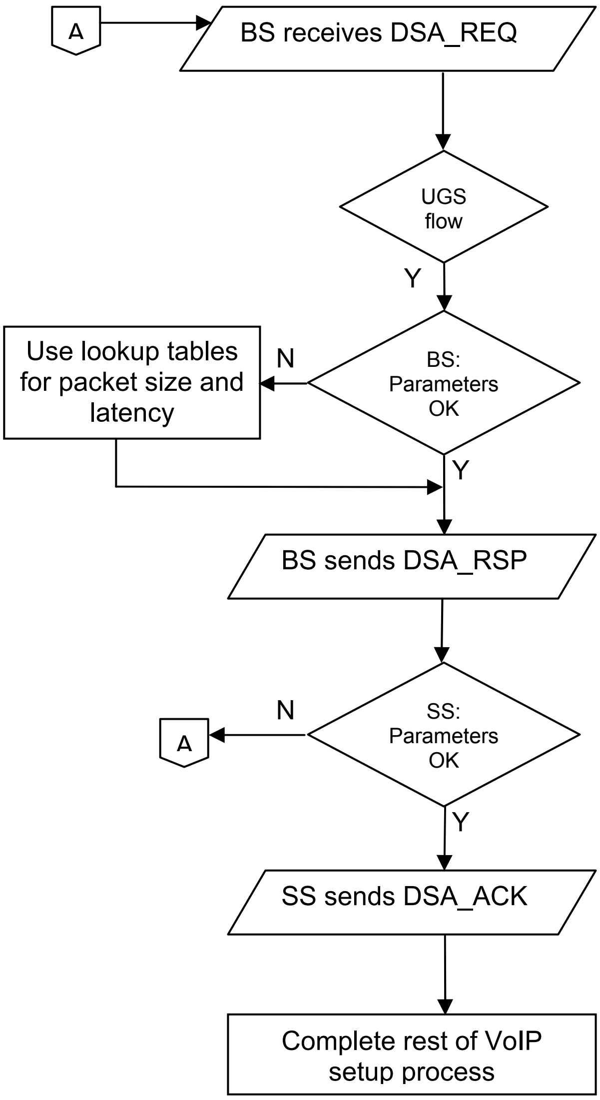


Fig. 4. Flow diagram of the procedure to determine an optimal parameter set at the start of a UGS service flow. In the first decision box if the flow is not a UGS type the treatment will be different and is not shown here.

If the BS agrees to all the parameters specified it will echo these back in a DSA_RSP message. (It would seem logical that this step occurs after the SS has setup the session with the receiver using H.323, SIP or another setup protocol) The procedure for this is shown in Fig. 4.

If however the requested parameters are not optimal and can be substituted by more efficient ones the BS will indicate these in the DSA_RSP message. Once a set of values is agreed upon the SS will confirm the use of the parameters by sending a DSA_ACK to the BS.

For applications which cannot change t_{pkt} the SS should indicate this to the BS. We propose using one of the unused Service Flow Parameters [1] in the DSA_REQ as an indicator. The BS will not attempt to optimize such parameters.

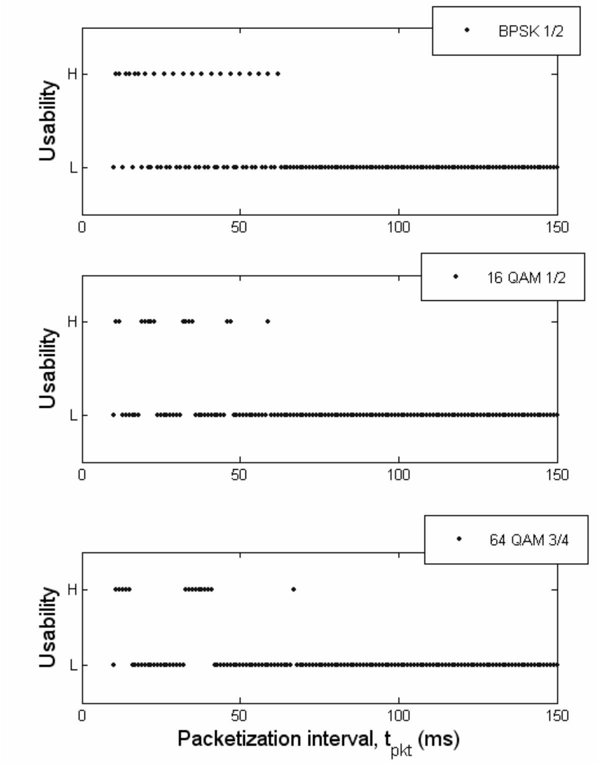


Fig. 5. Usability of various packetization intervals. H and L indicate High and Low usability respectively.

A. Lookup Table Creation and Usage

The BS needs to have lookup tables for a range of BER/SNR values and burst profiles, so that it can select the most suitable. Fig. 5 gives lookup table like data for BER of 1e-4. H and L indicate High and Low usability. As an example if a SS using 64 QAM 3/4 requests a 50ms interval the BS could respond with 1 of 2 possible options (lowest subplot of Fig. 5).

1) 67ms – the more efficient option. Has a higher latency and P_{loss}.

2) 41ms – less efficient that 1) but has lower latency and P_{loss}.

Based on channel conditions which the BS has knowledge of, and estimated delay to the destination it can select the best option.

It is important to remember that even though for the analysis we have considered t_{pkt} to be the unit of concern in Fig. 5, a lookup table at the BS would have to be based on packet size in Bytes. There is a linear relationship between them as given in (1).

V. Conclusion

The efficiency of bandwidth usage is affected by the choice of packet size in IEEE 802.16. This is more pronounced when the packet size is relatively small such as in VoIP applications. It has been shown that by careful selection of packetization intervals for VoIP the number of users can be increased and bandwidth wastage on overheads minimized. A modification was introduced to the MAC operation to be able to change the interval during call setup. This modification can be accommodated in the existing DSX hand shaking process so no extra overhead is introduced. Creating of a lookup table was proposed at the BS to make selecting an optimal interval fast and simple.

We are currently looking at the effect of packet size on the other scheduling types in 802.16.

References

[1] "IEEE standard for conformance to IEEE 802.16. Part 3: radio conformance tests (RCT) for 10-66 GHz WirelessMAN-SC air interface," in *IEEE Std 802.16/Conformance03-2004 (Conformance to IEEE Std 802.16-2001 as amended by IEEE Std 802.16a -2003 and IEEE Std 802.16c-2002)*, 2004.

[2] Y. Yao and J. Sun, "Study of UGS grant synchronization for 802.16," presented at Consumer Electronics, 2005. (ISCE 2005). Proceedings of the Ninth International Symposium on, 2005.

[3] H. Lee, T. Kwon, and D.-H. Cho, "An enhanced uplink scheduling algorithm based on voice activity for VoIP services in IEEE 802.16d/e system," *Communications Letters, IEEE*, vol. 9, pp. 691-693, 2005.

[4] H. Lee, T. Kwon, and D.-H. Cho, "An efficient uplink scheduling algorithm for VoIP services in IEEE 802.16 BWA systems," presented at Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th, 2004.

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

[6] G. Chu, D. Wang, and S. Mei, "A QoS architecture for the MAC protocol of IEEE 802.16 BWA system," presented at Communications, Circuits and Systems and West Sino Expositions, IEEE 2002 International Conference on, 2002.