

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 is performed in the context of the IEEE 802.16 MAC layer (specifically in the context of IEEE 802.16d, Fixed WiMAX). Bandwidth used for overheads which include lower layer headers as well as retransmissions at the MAC layer is considered. An optimal packetization interval selection method for real-time 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.

Keywords: *IEEE 802.16d, UGS, packetization interval, VoIP*

I. Introduction

THE need for mobile as well as fixed point wireless connectivity has increased greatly in recent years and new protocols and access technologies are emerging to satisfy this need. One of the most prominent 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 over metropolitan distances.

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. Amendment IEEE 802.16e (Mobile WiMAX) introduced extended rtPS (ertPS) to support VoIP with silence suppression. Even so there are several open 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 some 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 with "pure" UGS. Usage of rtPS and ertPS scheduling have also been investigated but they do 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 a few 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 overheads increases 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 physical layer parameters such as symbol size (in bits) when deciding the optimal packetization interval. We can safely assume the packet size is directly proportional to the packetization interval so we use this interval instead of packet size in the rest of the paper.

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, in section III we provide an analysis of the effect of packetization interval selection. We propose a method to dynamically select the best value during run time in section IV. A sample scenario is described in

section V followed by some simulation results in VI.

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 the 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) transmissions which are used for ranging, contention based bandwidth requests, allocated polled type bandwidth requests, DL PDUs and UL PDUs for SSs to send data to the BS. A MAP message could be compared to a table where the rows of the table are called information elements (IEs). Using these IE's the entire frame is explicitly defined in terms of start time, end time, who gets to transmit at what time and for how long. Ranging is a process which is usually 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. Ranging is even more important in the mobile scenario where adaptive modulation and coding (AMC) is used to keep error rates within acceptable limits.

The responsibility of scheduling UL/DL data belongs 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 three classes which can be used for real time flows, we will consider only UGS in this work although the results may 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) 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 contention request opportunities are prohibited for the UGS connection and only the periodic bandwidth grants maybe used. No piggyback requests or bandwidth stealing is allowed, but a SS can alert the BS to a bandwidth requirement of a non-UGS connection by using the poll me bit.

C. OFDM Physical Layer

IEEE 802.16 defines three possible physical layer configurations of which the orthogonal frequency division multiplexed (OFDM) is the best for non-line-of-sight access. We consider a 256 sub-carrier OFDM physical layer with the symbol sizes given in Table I. This table also gives the percentage of cell area covered by each modulation scheme.

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 symbol for each of them.

III. Analysis of Packetization Interval

In this section we analyze the effect the packetization interval has on system resource usage, packet loss rate, packet latency and the retransmit limit.

A. Packet Loss Rate

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.

$$n_{pkt} = r \times t_{pkt} + OH_{headers} \quad (1)$$

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

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

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 the number of symbols per packet.

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

$$PER = 1 - (1 - SER)^{n_{spp}} \quad (4)$$

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

$$P_{loss} = PER^{n+1} \quad (5)$$

B. Bandwidth Usage

$$OH_{ret} = n_{pdu} \sum_{i=1}^n \{i PER^i (1 - PER)\} + n_{pdu} \cdot n \cdot PER^{n+1} \quad (6)$$

$$OH_{bw} = \frac{(n_{pdu} - \text{payload}) \cdot (1 - PER^{n+1}) + OH_{ret}}{t_{pkt}} \quad (7)$$

Overhead due to retransmissions (in bits) is given in (6), where n is the MAC layer retransmit limit for this particular traffic class. The value of n has an upper bound which is defined in section III-D.

Bandwidth used for overhead is found by dividing the total overhead by the packetization interval, (7). The total BW requirement for the flow consists of the actual PDU and all subsequent retransmission averaged over t_{pkt} , (8).

E_f is the efficiency of the system and payload being the actual voice data from the application layer, (9).

$$BW = \frac{n_{pdu} + OH_{ret}}{t_{pkt}} \quad (8)$$

$$E_f = \frac{\text{payload} \cdot (1 - PER^{n+1})}{n_{pdu} + OH_{ret}} \quad (9)$$

C. Latency in Packet Delivery

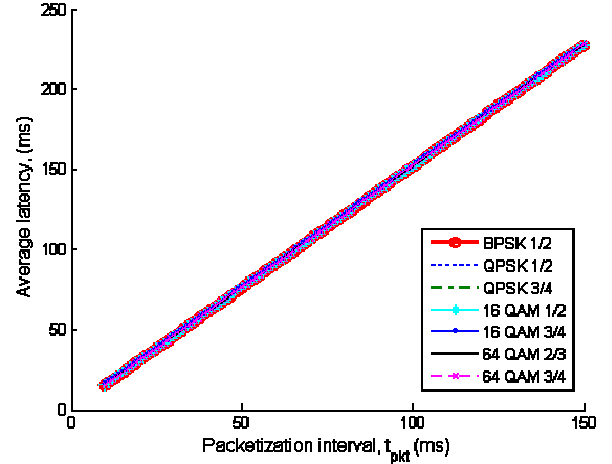


Fig. 1. Average latency for all the modulations schemes at a BER of $10^{-3.5}$. All the curves follow each other closely.

The average latency of a **successfully delivered** packet, l_{avg} is given by (10). This includes an additional component t_{pkt} in the summation which accounts for the lag due to packetization.

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

T_f is the frame duration. For low BER values ($<10^{-4}$) latency is roughly equal to $1.5t_{pkt}$. Even at higher BER's, since we are only considering the latency of successfully delivered packets, the average latency will only increase by a few T_f s. Figure 4 shows the latency for the different burst profiles. The minor difference is due to requiring multiple symbols for lower burst profiles.

Without latency minimizing enhancements we have considered the MAC service delay to be half of the packetization interval. At every talkspurt the starting point could randomly fall anywhere in the range $(0, t_{pkt}]$ from the current grant position.

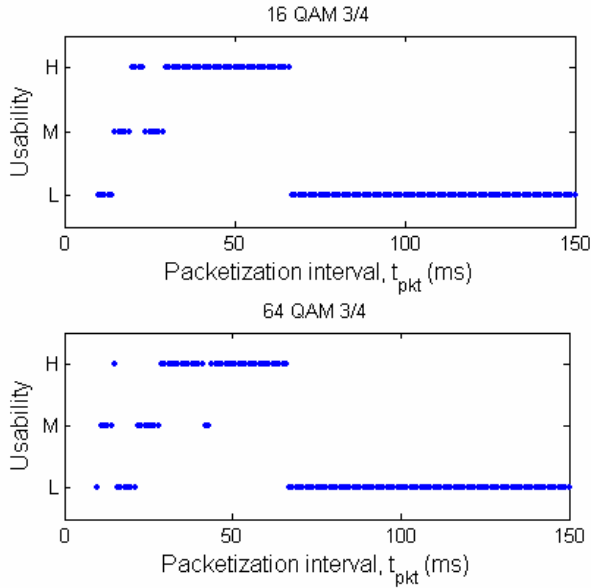


Fig. 2. Usability of various packetization intervals. H, M and L indicate High, Medium and Low usability respectively. BER is 10^{-5}

D. MAC Retransmit Limit

If a UL packet is corrupted beyond repair the BS will discard it and give the SS another BW allocation in the following frame. This method circumvents the need for the BS to provide explicit feedback for every packet. It also makes it possible to retransmit in consecutive frames and not alternating frames. By defining an Uplink Interface Usage Code (UIUC) specifically for retransmissions, the BS can implicitly request a retransmission of the last packet sent on a given Connection Identifier (CID). An IE will be included in the UL-MAP with the relevant CID and UIUC. If the SS has already refreshed its packet cache it will transmit a dummy packet with only a generic MAC header and no payload.

If the maximum number of MAC retransmissions based on BW constraints is defined as n_{max} , the retransmit limit is calculated as,

$$n = \min \left(\left\lfloor \frac{t_{pkt}}{T_f} \right\rfloor - 1, n_{max} \right) \quad (11)$$

A UGS grant will be given periodically so the maximum time available for retransmission is $t_{pkt} - T_f$. If it cannot be sent successfully within this time period the packet needs to be dropped.

IV. Proposed Implementation Scheme

A. Usability Factor, K

In order to fulfill the QoS requirements of the flow and operate within the BW constraints of the network, three conditions need to be satisfied.

$$\begin{aligned} P_{loss} &< P_{loss,max} \text{ (maximum } P_{loss}\text{)} \\ L_{avg} &< L_{max} \text{ (maximum tolerated latency)} \\ BW &< BW_{max} \text{ (maximum BW)} \end{aligned}$$

Losses cause a drop in VoIP quality and a drop rate greater than 1% is hard to conceal [7]. So $P_{loss,max}$ can be assumed to be 0.01.

Latency can cause voice quality degradation if it is excessive. The goal commonly used in designing networks to support VoIP is the target specified by ITU standard G.114. This states that 150 ms of one-way, end-to-end (from mouth to ear) delay ensures user satisfaction for telephony applications. This maximum delay should be apportioned to the various components of network delay (propagation delay through the backbone, scheduling delay because of congestion, and access link serialization delay) and service delay (because of VoIP gateway codec and de-jitter buffer). However, using a UGS flow the maximum latency will be bounded as explained above. When communicating with another SS in the same cell or in a cell which is part of the WIMAX network, the latency will be lower than when using the public internet, for example, international VoIP calls.

The BW_{max} which can be allocated to a flow very much depends on the SLA with the SS and also on the traffic mix and load dynamics.

We define a metric called Usability Factor, K , which combines all three metrics to give a value between 0 and 1. The closer it is to 1 the more suitable the t_{pkt} . A value of 0 implies that the interval cannot satisfy the three requirements, and should not be used.

$$K = S \times abs \left[\frac{(P_{loss,max} - P_{loss})(L_{max} - L_{avg})(BW_{max} - BW)}{(P_{loss,max} - 0)(L_{max} - 1.5t_{pkt})(BW_{max} - r)} \right] \quad (12)$$

$$S = \begin{cases} 1 & , (P_{loss,max} > P_{loss}) \& (L_{max} > L_{avg}) \\ & \& (BW_{max} > BW) \\ 0 & , (P_{loss,max} < P_{loss}) \text{ or } (L_{max} < L_{avg}) \\ & \text{ or } (BW_{max} < BW) \end{cases} \quad (13)$$

As stated before, r is the nominal bitrate of the VoIP application. The denominator of (12) is the product of the three constraints less the best

attainable conditions. abs is the absolute value and S is given in (13) above. S has the effect of seeding out only areas of the function where all three conditions are met. Any regions where at least one of the conditions is not true are set to zero.

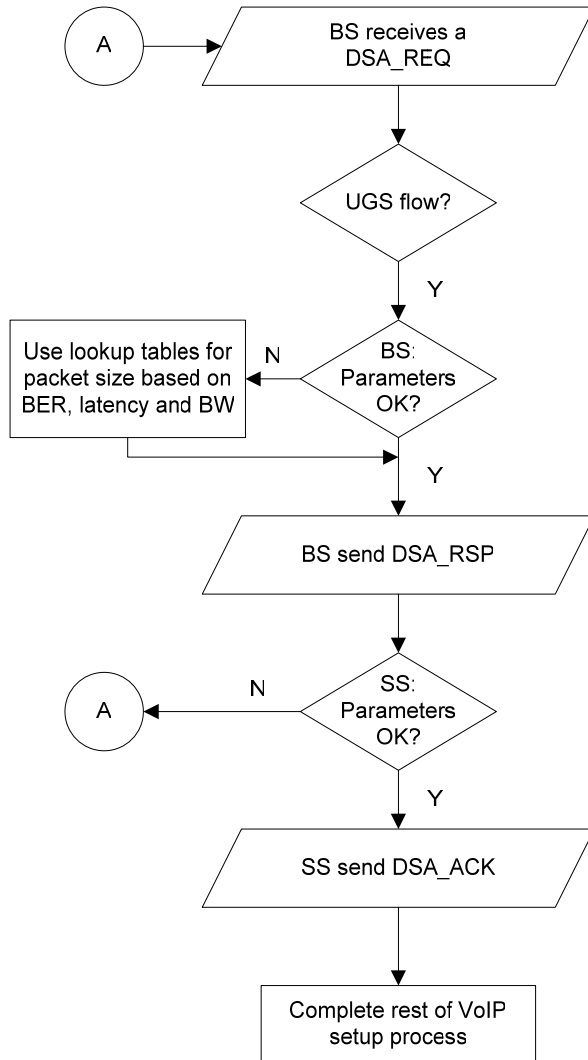


Fig. 3 Procedure to determine an optimal parameter set at the start of a UGS service flow. In the first decision box if the flow is not a UGS type the treatment will be different and is not shown here.

B. 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. The usability factor can be directly used for creating the lookup tables or it can be quantized to, say, three levels such as done in Figure 2. (The usability values shown in Figures 6, 7 and 8 relating to the sample scenario of section V, could be quantized to the form of

Figure 2)

We use High (H), Medium (M) and Low (L). L implies one or more of the constraints will not be satisfied. This makes the lookup table simple and less memory is needed to store it. As an example, if a SS using 64 QAM 3/4 requests a 20 ms interval the BS could respond with anything between 29 ms and 41 ms.

- 1) 41 ms – the more efficient option. Has a higher latency and P_{loss} , though.
- 2) 29 ms – less efficient than 1) but has lower latency and P_{loss} .

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

C. DSX Process for Setting/Updating t_{pkt}

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.

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 before the SS has setup the session with the receiver using H.323, SIP or another setup protocol) The procedure for this is shown in Figure 3.

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.

In the mobile scenario, or when channel conditions vary with time, the SS will carry out periodic ranging to update its modulation scheme. Every time an update takes place the BS will initiate a Dynamic Service Change (DSC) handshake process to renegotiate the parameters of the VoIP session. The MAC then informs the application of the required changes, which in turn, will inform the other end-point of the conversation.

D. Increase in the Number of Users

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 of these

annuluses mark the change to a lower modulation scheme. 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 only the three lowest modulation schemes. Table I lists the possible modulation schemes, bitrates and proportion of annuluses. BPSK 2/3 is not included in the table because it has a very similar SNR requirement to BPSK 1/2. We assume that the total number of SS's in the cell is uniformly spread out. $b_{r,i}$ is the effective bit rate of a SS in the i^{th} annulus for a randomly chosen t_{pkt} . $b_{o,i}$ is the effective bit rate of a SS in the i^{th} annulus for an optimally chosen t_{pkt} . n_r and n_o are the numbers of users in the system for random t_{pkt} and optimal t_{pkt} , using the same amount of resources measured in OFDM symbols. Equation (14) gives the proportional increase in the number of users due to better selection of t_{pkt} . Here m_l and m_h represent the index of the lowest and highest modulation schemes respectively. This increase is plotted in Figure 4.

$$\frac{n_o}{n_r} = \frac{\sum_{i=m_l}^{m_h} n_r a_i \left(\frac{b_{r,i}}{n_{bps,i}} \right)}{\sum_{i=m_l}^{m_h} n_o a_i \left(\frac{b_{o,i}}{n_{bps,i}} \right)} = \frac{\sum_{i=m_l}^{m_h} a_i \left(\frac{b_{r,i}}{n_{bps,i}} \right)}{\sum_{i=m_l}^{m_h} a_i \left(\frac{b_{o,i}}{n_{bps,i}} \right)} \quad (14)$$

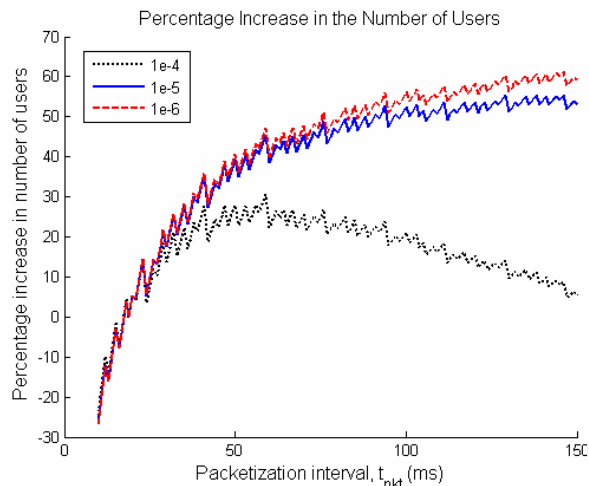


Fig. 4. Percentage increase in the number of users for a fixed amount of UL resources. Three values of BER are compared.

The cell boundary of an ideal cell in a cellular

infrastructure is a hexagon so the boundary of the lowest modulation scheme is considered to be hexagonal. **We also assume the cell is large enough so that the cell boundary is in effect the transmission range of the lowest modulation scheme.**

V. Sample Scenario

A sample scenario using common values for 802.16 is now considered to demonstrate the effects of t_{pkt} . The analysis was carried out using Matlab. 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 with a TDD frame structure and a 4ms frame duration. 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 data stream at a rate of 32 kbps. This is an implementation dependant value. We have used 32 kbps as it will produce a good quality voice stream [7]. The retransmission limit is set as per (11). The maximum number of tolerated residual bit errors in a packet reaching the MAC layer is 0.

Modulation Scheme	Bits per OFDM symbol (n_{bps})	Percentage of total area (a_i)
16 QAM 1/2	384	20.1
16 QAM 3/4	576	51.5
64 QAM 1/2	768	9.2
64 QAM 3/4	864	19.2

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

We consider a cell with a radius of 2 km. Using the SNR thresholds in Table I, it can be shown that the entire cell can be covered using the three highest modulation schemes. Nevertheless we include the next lower modulation scheme to cope with shadowing variations. A two-ray ground propagation model was used to calculate receiver SNRs. The revised coverage percentages are given in Table II above.

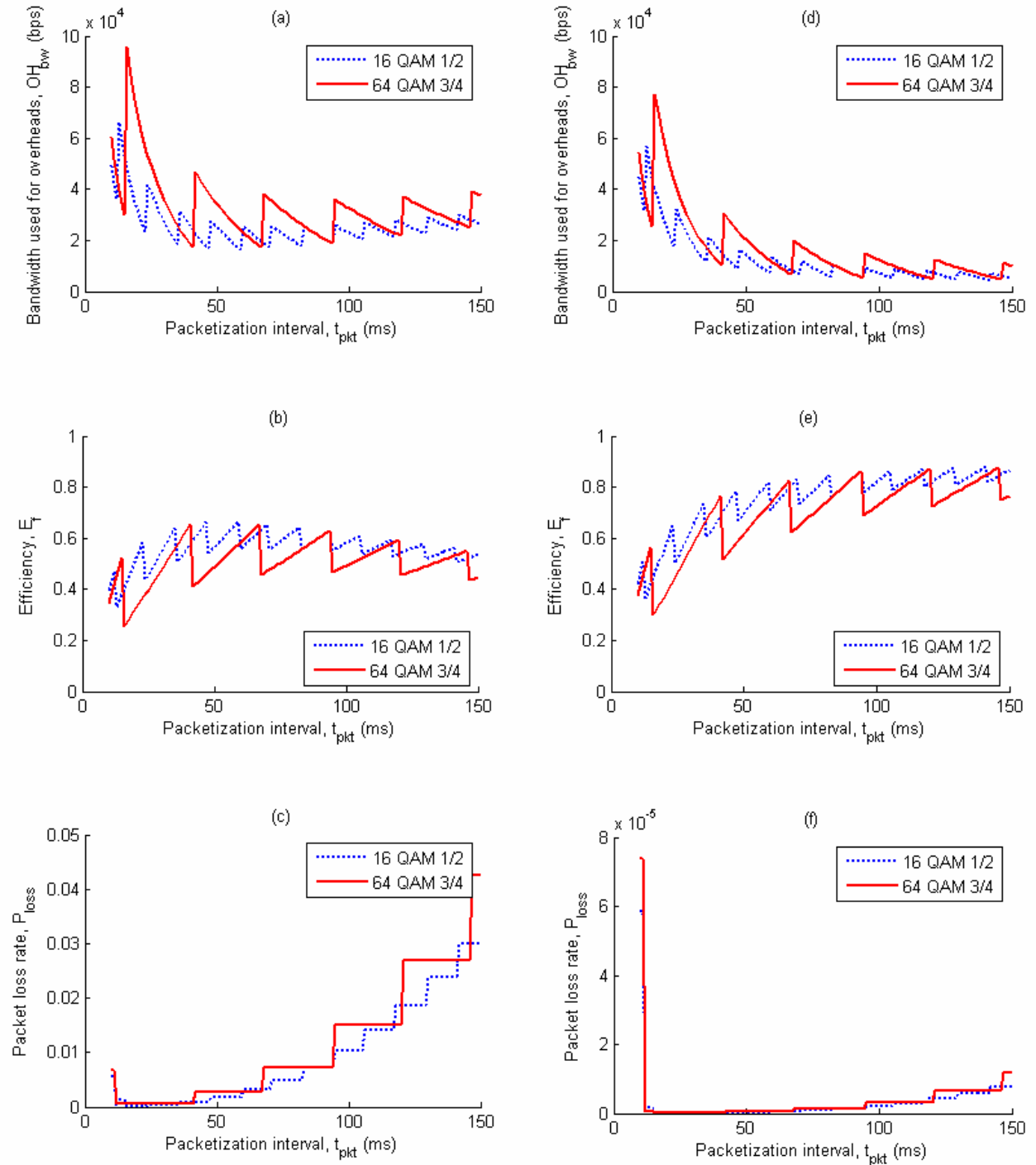


Fig. 5. (a)-(c) are plots of overhead bandwidth, efficiency and PER for a BER of 10^{-4} and (d)-(f) are for a BER of 10^{-5} . The two modulation schemes shown here are 16QAM $\frac{1}{2}$ and 64QAM $\frac{3}{4}$. The spiky nature of the overhead bandwidth and efficiency plots as well as the staircase shape of the PER plots is due to the packet size being an integer number of OFDM symbols.

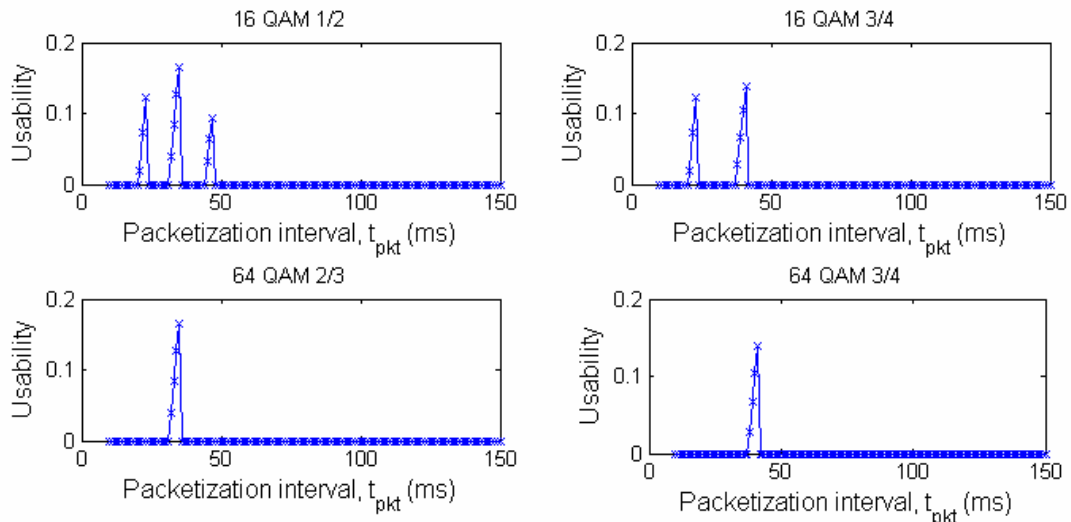


Fig. 6. Usability factor, K , for a BER of $10^{-3.5}$.

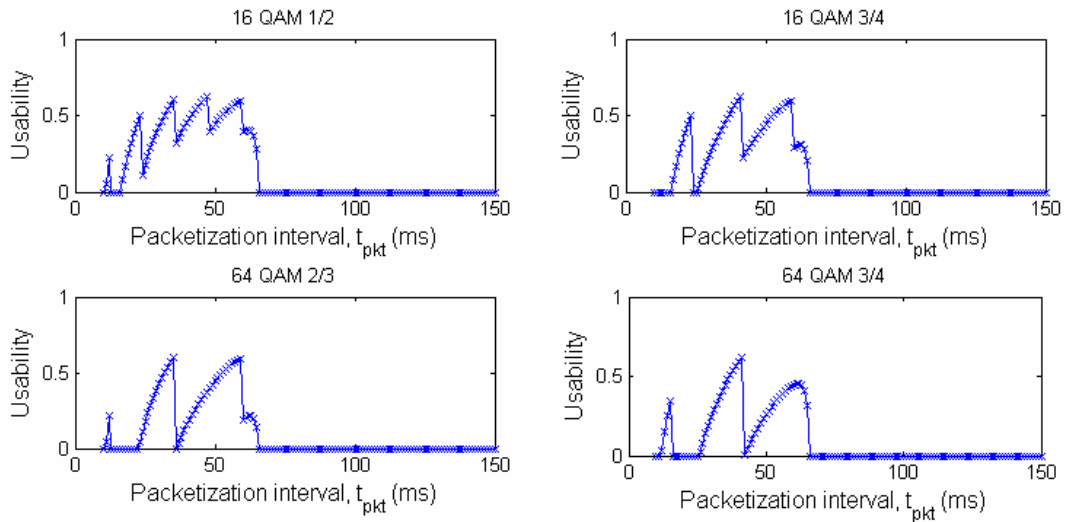


Fig. 7. Usability factor, K , for a BER of 10^{-4} .

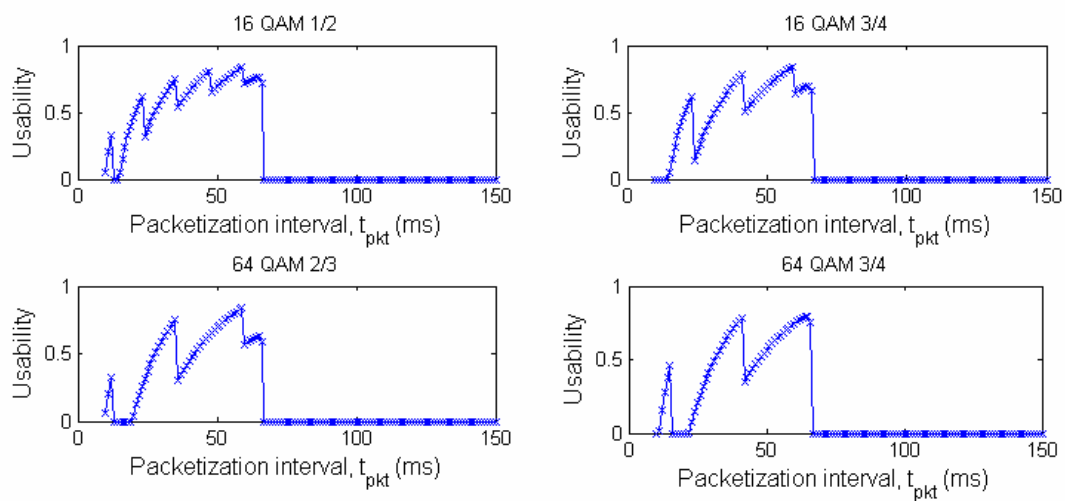


Fig. 8. Usability factor, K , for a BER of 10^{-5} .

Figure 1 gives the Efficiency, overhead bandwidth and packet loss rate for packetization intervals ranging from 10ms to 150ms with BER's of 10^{-4} and 10^{-5} . The saw-tooth effect is due to the transmission units being integer multiples of OFDM symbols.

When the BER is lower as in Figure 1(d)-(f) 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.

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

The higher P_{loss} values at small intervals, Figures 1(c) and 1(f) can be explained by considering (11). At small intervals, the retransmit limit is also low which reduces the probability of recovering erroneously received packets.

In order to select the most appropriate interval, we have plotted the usability factor, K in Figures 6, 7 and 8 for BER's of $10^{-3.5}$, 10^{-4} and 10^{-5} , respectively. The constraints used are,

$$\begin{aligned} P_{loss,max} &= 0.01 \\ L_{max} &= 100 \text{ ms} \\ BW_{max} &= 80 \text{ kbps} \end{aligned}$$

At a BER of $10^{-3.5}$ only a few intervals are usable. (When the BER is 10^{-3} none of intervals can be used under the given conditions so this has not been plotted.) These intervals only barely satisfy the three constraints. Constraints 1 and 3 have the biggest effect at high BER's.

At lower BER's the choice increases and more intervals will satisfy the requirements as seen in Figures 7 and 8. Constraint 2 (L_{max}) limits the maximum value of the interval while constraint 3 (BW_{max}) limits the minimum value of the interval. So constraints 2 and 3 dominate. The closer K is to 1, the better the interval.

The ratio of $n_o:n_r$ given in (14) is converted to a percentage increase in Figure 9. Here we have used 20 ms t_{pkt} as the reference. The negative portion of the curves represents a reduction in the total users relative to the 20 ms case. If we compare Figures 4 and 9, it can be seen that the increase is more in this scenario where only the highest four burst profiles are used. The curves are also more jagged. At higher burst profiles, OFDM symbols consist of a large number of bits. For most t_{pkt} 's the packet can be represented by a small number of symbols.

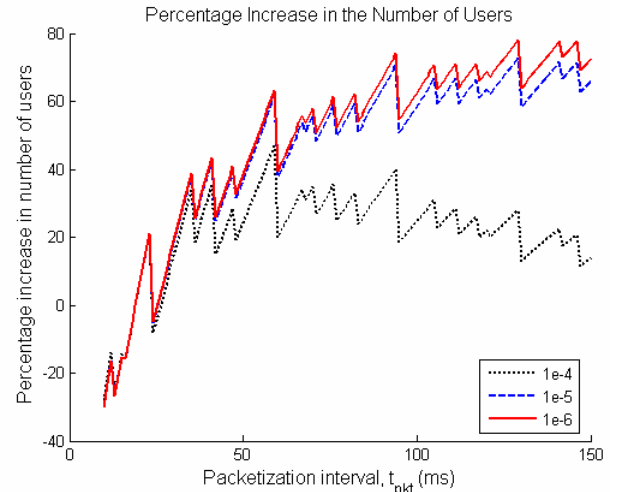


Fig. 9. Percentage increase in the number of users for a fixed amount of UL resources. Three values of BER are compared.

Poor choices of t_{pkt} can cause almost a whole symbol to be unused which can have a significant effect on the total cell throughput. This in turn can cause the number of sustainable users to vary significantly for a small change in t_{pkt} . It is also important to note that in this scenario at a BER of 10^{-6} only t_{pkt} values up to 67 ms need be looked at due to the latency constraint, Figure 8. Similarly for the other BER value the upper bounds for t_{pkt} should be taken from Figures 6 and 7. Based on Equation (10), 67 ms would give a minimum latency of 100 ms.

VI. Simulation Results

An IEEE 802.16d model has been implemented using QualNet3.8 [8]. A single mobile SS is initialized close to the BS location and it then moves at constant speed towards the cell boundary. A cell radius of 3 km is used and time to traverse this distance is 80 s. These are arbitrary values chosen to easily observe the periodic ranging process and step down of modulation schemes within a short simulation time. Periodic ranging is carried out every 10 s which is adequate for this relatively high speed mobility scenario. In addition ranging can also be done at dynamic intervals whenever the SS detects a drop or rise in SNR.

A 32 kbps Constant Bit Rate (CBR) packet flow from the SS to the BS is used to simulate the UL stream of a VoIP session. The CBR application was modified in order to accommodate dynamic packet size and interval adjustments as and when indicated by the MAC layer. When the MAC detects a drop in the SNR (or increase if moving toward the BS) it will initiate ranging. The BS on

receiving the RNG-REQ checks if any VoIP connections are active for that SS. If so it will initiate a DSC process as soon as ranging is completed. In the simulation the SS accepts the recommended parameters without requiring further negotiations.

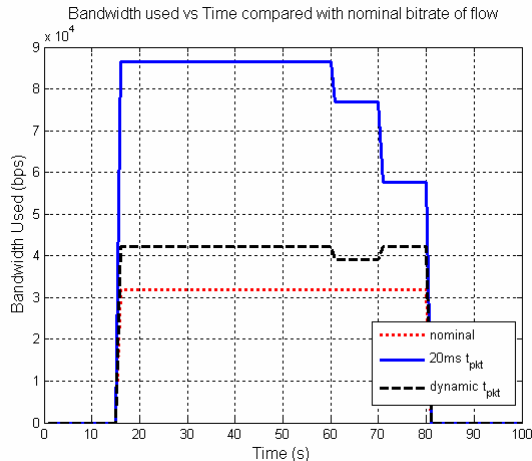


Fig. 10. Comparison of bandwidth used for 20 ms t_{pkt} and dynamic t_{pkt} . The nominal bitrate of the flow (32 kbps) is also shown.

Initially the application starts with a t_{pkt} of 20 ms but this is increased to 41 ms during the DSA handshake. It is assumed that the SS will accept the parameters proposed by the BS. The 3-way handshake during periodic ranging is also simplified to a 2-way handshake for ease of modeling. An error free link is considered.

As the SS moves away from the BS, until 60 s it is able to use 64QAM 3/4. At 60 s, after periodic ranging it steps down to the next lower modulation scheme, 64QAM 2/3. At 70 s, another step down occurs to 16QAM 3/4. The t_{pkt} values used are taken from Figure 8, since K values for 0 BER are very similar to those for 10^{-5} BER. Actually usable K values for a higher BER are a subset of those usable at a lower BER. Figure 11 clearly shows the lower bandwidth usage due to the dynamic selection of t_{pkt} .

We stress that fixed WiMAX cannot support the mobility scenario given in the above simulation but is used simply to observe the response of the modified MAC layer, to adaptive modulation and coding (AMC) in a short time span.

VII. Conclusion

The efficiency of bandwidth usage, packet loss rate and latency are 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 was shown that by careful selection of the packetization intervals for VoIP the number of users can be increased and bandwidth wastage on overheads minimized. A selection criterion based on three constraints was introduced. A modification was proposed to the MAC operation to be able to change the interval during call setup and also during periodic ranging. This modification can be accommodated in the existing DSX handshaking process, so no extra overhead is introduced. Creation of a lookup table at the BS was proposed to make selecting an optimal interval simple and fast. A proof of concept was also given based on a simple simulation scenario.

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.
- [7] C. Hattingh and T. Szigeti, *End-to-End QoS Network Design: Quality of Service in LANs, WANs, and VPNs*, 1 ed: Cisco Press., 2004.
- [8] <http://www.qualnet.com/>.



Shehan Perera (BE 2002, Canterbury, New Zealand) was born in Colombo, Sri Lanka. He completed his primary and secondary education in Colombo. In 2002 he received a bachelor's degree with first class honours in Electrical and Electronics Engineering and has since been working towards a PhD at the University of Canterbury. His area of research is QoS in wireless broadband networks with a special interest in WiMAX technology.



Professor Harsha Sirisena (BSc (Eng) Electrical Engineering, Ceylon, PhD Control Engineering, Cambridge) has been with the University of Canterbury since 1971 where he is currently a Professor in the Department of Electrical and Computer Engineering and leads the Networking research group.

He has held visiting academic positions at the Universities of Calgary, Lund and Minnesota, the Australian National University, the Hong Kong Polytechnic University and the National University of Singapore.

His areas of interest include performance and quality of service issues in next generation wireline and wireless networks including WiMAX.



Professor Krys Pawlikowski (Ph.D. (Distinction) Computer Engineering, Technical University of Gdansk, Poland) has been an academic member of staff in the Department of Computer Science at the University of Canterbury in Christchurch, New Zealand, since 1984.

He has given many invited lectures and seminars at universities and research institutions in Asia, Australia, Europe and North America, as well as invited talks at international conferences around the world.

He is the author or co-author of more than 130 research publications, including 4 books on computer networks; vice-chairman and member of numerous program committees of international conferences.

At the University of Canterbury, he has been co-founder of the Network Research Centre on Network Architectures and Protocols, Distributed Processing and Simulation and leader of a research project AKAROA on distributed stochastic simulation.

His research interests include: performance modeling of telecommunications networks, teletraffic modeling, distributed stochastic simulation and simulation output data analysis.