

Research Article

Capacity Evaluation for IEEE 802.16e Mobile WiMAX

Chakchai So-In, Raj Jain, and Abdel-Karim Tamimi

Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

Correspondence should be addressed to Chakchai So-In, cs5@cse.wustl.edu

Received 21 September 2009; Accepted 2 December 2009

Academic Editor: Rashid Saeed

Copyright © 2010 Chakchai So-In et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a simple analytical method for capacity evaluation of IEEE 802.16e Mobile WiMAX networks. Various overheads that impact the capacity are explained and methods to reduce these overheads are also presented. The advantage of a simple model is that the effect of each decision and sensitivity to various parameters can be seen easily. We illustrate the model by estimating the capacity for three sample applications—Mobile TV, VoIP, and data. The analysis process helps explain various features of IEEE 802.16e Mobile WiMAX. It is shown that proper use of overhead reducing mechanisms and proper scheduling can make an order of magnitude difference in performance. This capacity evaluation method can also be used for validation of simulation models.

1. Introduction

IEEE 802.16e Mobile WiMAX is the standard [1] for broadband (high-speed) wireless access (BWA) in a metropolitan area. Many carriers all over the world have been deploying Mobile WiMAX infrastructure and equipment. For interoperability testing, several WiMAX profiles have been developed by WiMAX Forum.

The key concern of these providers is how many users they can support for various types of applications in a given environment or what value should be used for various parameters. This often requires detailed simulations and can be time consuming. In addition, studying sensitivity of the results to various input values requires multiple runs of the simulation further increasing the cost and complexity of the analysis. Therefore, in this paper we present a simple analytical method of estimating the number of users on a Mobile WiMAX system. This model has been developed for and used extensively in WiMAX Forum [2].

There are four goals of this paper. First, we want to present a simple way to compute the number of users supported for various applications. The input parameters can be easily changed allowing service providers and users to see the effect of parameter change and to study the sensitivity to various parameters. Second, we explain all the factors that affect the performance. In particular, there

are several overheads. Unless steps are taken to avoid these, the performance results can be very misleading. Note that the standard specifies these overhead reduction methods; however, they are not often modeled. Third, proper scheduling can make an order of magnitude difference in the capacity since it can change the number of bursts and the associated overheads significantly. Fourth, the method can also be used to validate simulation models that can handle more sophisticated configurations.

This paper is organized as follows. In Section 2, we present an overview of Mobile WiMAX physical layer (PHY). Understanding this is important for performance modeling. In Section 3, Mobile WiMAX system and configuration parameters are discussed. The key input to any capacity planning and evaluation exercise is the workload. We present three sample workloads consisting of Mobile TV, VoIP, and data applications in Section 4. Our analysis is general and can be used for any other application workload. Section 5 explains both upper and lower layer overheads and ways to reduce those overheads. The number of users supported for the three workloads is finally presented in Section 6. It is shown that with proper scheduling, capacity can be improved significantly. Both error-free perfect channel and imperfect channel results are also presented. Finally, the conclusions are drawn in Section 7.

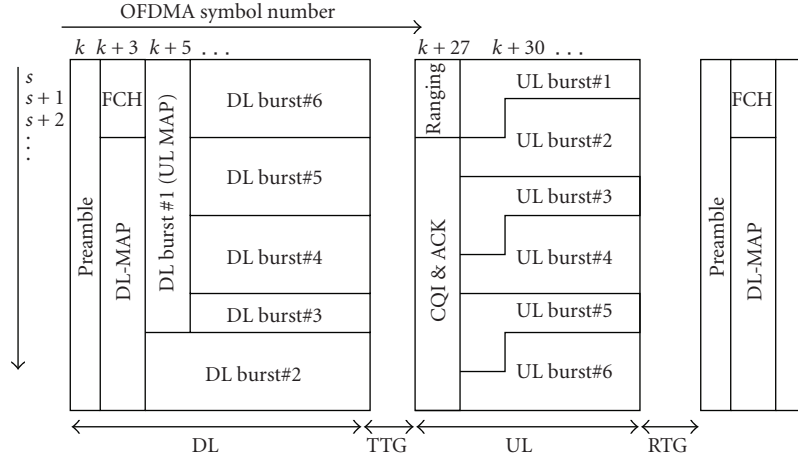


FIGURE 1: A Sample OFDMA frame structure.

2. Overview of Mobile WiMAX PHY

One of the key developments of the last decade in the field of wireless broadband is the practical adoption and cost effective implementation of an Orthogonal Frequency Division Multiple Access (OFDMA). Today, almost all upcoming broadband access technologies including Mobile WiMAX and its competitors use OFDMA. For performance modeling of Mobile WiMAX, it is important to understand OFDMA. Therefore, we provide a very brief explanation that helps us introduce the terms that are used later in our analysis. For further details, we refer the reader to one of several good books and survey on Mobile WiMAX [3–7].

Unlike WiFi and many cellular technologies which use fixed width channels, Mobile WiMAX allows almost any available spectrum width to be used. Allowed channel bandwidths vary from 1.25 MHz to 28 MHz. The channel is divided into many equally spaced subcarriers. For example, a 10 MHz channel is divided into 1024 subcarriers some of which are used for data transmission while others are reserved for monitoring the quality of the channel (pilot subcarriers), for providing safety zone (guard subcarriers) between the channels, or for using as a reference frequency (DC subcarrier).

The data and pilot subcarriers are modulated using one of several available MCSs (Modulation and Coding Schemes). Quadrature Phase Shift Keying (QPSK) and Quadrature Amplitude Modulation (QAM) are examples of modulation methods. Coding refers to the Forward Error Correction (FEC) bits. Thus, QAM-64 1/3 indicates an MCS with 6-bit (64 combinations) QAM modulated symbols and the error correction bits take up 2/3 of the bits leaving only 1/3 for data.

In traditional cellular networks, the downlink—Base Station (BS) to Mobile Station (MS)—and uplink (MS to BS) use different frequencies. This is called Frequency Division Duplexing (FDD). Mobile WiMAX allows not only FDD but also Time Division Duplexing (TDD) in which the downlink (DL) and uplink (UL) share the same frequency but alternate

in time. The transmission consists of frames as shown in Figure 1. The DL subframe and UL subframe are separated by a TTG (Transmit to Transmit Gap) and RTG (Receive to Transmit Gap). The frames are shown in two dimensions with frequency along the vertical axis and time along the horizontal axis.

In OFDMA, each MS is allocated only a subset of the subcarriers. The available subcarriers are grouped into a few subchannels and the MS is allocated one or more subchannels for a specified number of symbols. The mapping process from logical subchannel to multiple physical subcarriers is called a permutation. Basically, there are two types of permutations: distributed and adjacent. The distributed subcarrier permutation is suitable for mobile users while adjacent permutation is for fixed (stationary) users. Of these, Partially Used Subchannelization (PUSC) is the most common used in a mobile wireless environment [3]. Others include Fully Used Subchannelization (FUSC) and Adaptive Modulation and Coding (band-AMC). In PUSC, subcarriers forming a subchannel are selected randomly from all available subcarriers. Thus, the subcarriers forming a subchannel may not be adjacent in frequency.

Users are allocated a variable number of *slots* in the downlink and uplink. The exact definition of slots depends upon the subchannelization method and on the direction of transmission (DL or UL). Figures 2 and 3 show slot formation for PUSC. In uplink (Figure 2), a slot consists of 6 *tiles* where each tile consists of 4 subcarriers over 3 symbol times. Of the 12 subcarrier-symbol combinations in a tile, 4 are used for pilot and 8 are used for data. The slot, therefore, consists of 24 subcarriers over 3 symbol times. The 24 subcarriers form a subchannel. Therefore, at 10 MHz, 1024 subcarriers form 35 UL subchannels. The slot formation in downlink is different and is shown in Figure 3. In the downlink, a slot consists of 2 clusters where each cluster consists of 14 subcarriers over 2 symbol times. Thus, a slot consists of 28 subcarriers over two symbol times. The group of 28 subcarriers is called a subchannel resulting in 30 DL subchannels from 1024 subcarriers at 10 MHz.

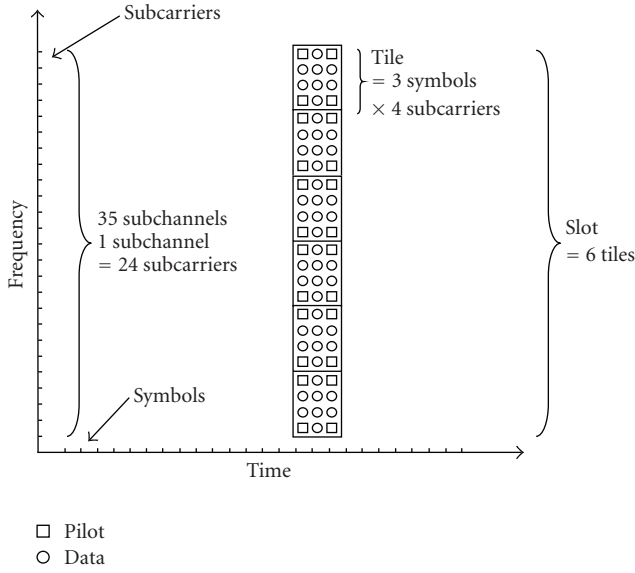


FIGURE 2: Symbols, tiles, and slots in uplink PUSC.

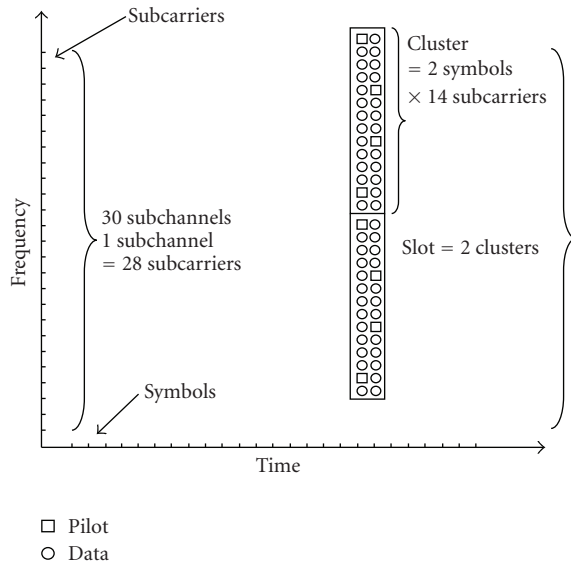


FIGURE 3: Symbols, clusters, and slots in downlink PUSC.

The Mobile WiMAX DL subframe, as shown in Figure 1, starts with one symbol-column of preamble. Other than preamble, all other transmissions use slots as discussed above. The first field in DL subframe after the preamble is a 24-bit Frame Control Header (FCH). For high reliability, FCH is transmitted with the most robust MCS (QPSK 1/2) and is repeated 4 times. Next field is DL-MAP which specifies the burst profile of all user bursts in the DL subframe. DL-MAP has a fixed part which is always transmitted and a variable part which depends upon the number of bursts in DL subframe. This is followed by UL-MAP which specifies the burst profile for all bursts in the UL subframe. It also consists of a fixed part and a variable part. Both DL MAP and UL MAP are transmitted using QPSK 1/2 MCS.

3. Mobile WiMAX Configuration Parameters and Characteristics

The key parameters of Mobile WiMAX PHY are summarized in Tables 1 through 3.

Table 1 lists the OFDMA parameters for various channel widths. Note that the product of subcarrier spacing and FFT size is equal to the product of channel bandwidth and sampling factor. For example, for a 10 MHz channel, $10.93 \text{ kHz} \times 1024 = 10 \text{ MHz} \times 28/25$. This table shows that at 10 MHz the OFDMA symbol time is 102.8 microseconds and so there are 48.6 symbols in a 5 millisecond frame. Of these, 1.6 symbols are used for TTG and RTG leaving 47 symbols. If n of these are used for DL, then $47 - n$ are available for uplink. Since DL slots occupy 2 symbols and UL slots occupy 3 symbols, it is best to divide these 47 symbols such that $47 - n$ is a multiple of 3 and n is of the form $2k + 1$. For a DL : UL ratio of 2 : 1, these considerations would result in a DL subframe of 29 symbols and UL subframe of 18 symbols. In this case, the DL subframe will consist of a total of 14×30 or 420 slots. The UL subframe will consist of 6×35 or 210 slots.

Table 2 lists the number of data, pilot, and guard subcarriers for various channel widths. A PUSC subchannelization is assumed, which is the most common subchannelization [3].

Table 3 lists the number of bytes per slot for various MCS values. For each MCS, the number of bytes is equal to $[\text{number bits per symbols} \times \text{Coding Rate} \times 48 \text{ data subcarriers and symbols per slot}/8 \text{ bits}]$. Note that for UL, the maximum MCS level is QAM-16 2/3 [2].

This analysis method can be used for any allowed channel width, any frame duration, or any subchannelization. We assume a 10 MHz Mobile WiMAX TDD system with 5-millisecond frame duration, PUSC subchannelization mode, and a DL : UL ratio of 2 : 1. These are the default values recommended by Mobile WiMAX forum system evaluation methodology and are also common values used in practice. The number of DL and UL slots for this configuration can be computed as shown in Table 4.

4. Traffic Models and Workload Characteristics

The key input to any capacity planning exercise is the workload. In particular, all statements about number of subscribers supported assume a certain workload for the subscriber. The main problem is that workload varies widely with types of users, types of applications, and time of the day. One advantage of the simple analytical approach presented in this paper is that the workload can be easily changed and the effect of various parameters can be seen almost instantaneously. With simulation models, every change would require several hours of simulation reruns. In this section, we present three sample workloads consisting of Mobile TV, VoIP, and data applications. We use these workloads to demonstrate various steps in capacity estimation.

The VoIP workload is symmetric in that the DL data rate is equal to the UL data rate. It consists of very small packets that are generated periodically. The packet size and

TABLE 1: OFDMA parameters for Mobile WiMAX [3, 8, 9].

Parameters	Values						
System bandwidth (MHz)	1.25	5	10	20	3.5	7	8.75
Sampling factor	28/25				8/7		
Sampling frequency (F_s , MHz)	1.4	5.6	11.2	22.4	4	8	10
Sample time ($1/F_s$, nsec)	714	178	89	44	250	125	100
FFT size (N_{FFT})	128	512	1,024	2,048	512	1,024	1,024
Subcarrier spacing (Δf , kHz)	10.93				7.81		9.76
Useful symbol time ($T_b = 1/\Delta f$, μs)	91.4				128		102.4
Guard time ($T_g = T_b/8$, μs)	11.4				16		12.8
OFDMA symbol time ($T_s = T_b + T_g$, μs)	102.8				144		115.2

TABLE 2: Number of subcarriers in PUSC [8].

Parameters			Values		
(a) DL					
System bandwidth (MHz)	1.25	2.5	5	10	20
FFT size	128	N/A	512	1,024	2,084
number of guard subcarriers	43	N/A	91	183	367
number of used subcarriers	85	N/A	421	841	1,681
number of pilot subcarriers	12	N/A	60	120	240
number of data subcarriers	72	N/A	360	720	140
(b) UL					
System bandwidth (MHz)	1.25	2.5	5	10	20
FFT size	128	N/A	512	1,024	2,084
number of guard subcarriers	31	N/A	103	183	367
number of used subcarriers	97	N/A	409	841	1,681

TABLE 3: Slot capacity for various MCSs.

MCS	Bits per symbol	Coding Rate	DL bytes per slot	UL bytes per slot
QPSK 1/8	2	0.125	1.5	1.5
QPSK 1/4	2	0.25	3.0	3.0
QPSK 1/2	2	0.50	6.0	6.0
QPSK 3/4	2	0.75	9.0	9.0
QAM-16 1/2	4	0.50	12.0	12.0
QAM-16 2/3	4	0.67	16.0	16.0
QAM-16 3/4	4	0.75	18.0	16.0
QAM-64 1/2	6	0.60	18.0	16.0
QAM-64 2/3	6	0.67	24.0	16.0
QAM-64 3/4	6	0.75	27.0	N/A
QAM-64 5/6	6	0.83	30.0	N/A

the period depend upon the vocoder used. G723.1 Annex A is used in our analysis and results in a data rate of 5.3 kbps, 20 bytes voice packet every 30 millisecond. Note that other vocoder parameters can be also used and they are listed in Table 5.

The Mobile TV workload depends upon the quality and size of the display. In our analysis, a sample measurement on a small screen Mobile TV device produced an average packet size of 984 bytes every 30 millisecond resulting in an

average data rate of 350.4 kbps [11, 12]. Note that Mobile TV workload is highly asymmetric with almost all of the traffic going downlink. Table 6 also shows other types of Mobile TV workload.

For data workload, we selected the Hypertext Transfer Protocol (HTTP) workload recommended by the 3rd Generation Partnership Project (3GPP) [13]. The parameters of HTTP workload are summarized in Table 7.

The characteristics of the three workloads are summarized in Table 8. In this table, we also include higher level headers, that is, IP, UDP, and TCP, with a header compression mechanism. Detailed explanation of PHS (Payload Header Suppression) and ROHC (Robust Header Compression) is presented in the next section. Given ROHC, the data rate with higher level headers ($R_{\text{with Header}}$) is calculated by

$$R_{\text{with Header}} = R \times \frac{(\text{MSDU} + \text{Header})}{\text{MSDU}}. \quad (1)$$

Here, MSDU is the MAC SDU size and R is the application data rate. Given the R , number of bytes per frame per user can be derived from $R_{\text{with Header}} \times \text{frame.duration}$. For example, for Mobile TV, with 983.5 bytes of MAC SDU size and 350 kbps of application data rate, with ROHC type 1, MAC SDU size with header is $983.5 + 1$ bytes and as a result, the data rate with header is 350.4 kbps and results in 216 bytes per frame.

TABLE 4: Mobile WiMAX system configurations.

Configurations	Downlink	Uplink
DL and UL symbols excluding preamble	28	18
Ranging, CQI, and ACK (symbols columns)	N/A	3
number of symbol columns per Cluster/Tile	2	3
number of subcarriers per Cluster/Tile	14	4
Symbols \times Subcarriers per Cluster/Tile	28	12
Symbols \times Data Subcarriers per Cluster/Tile	24	8
number of pilots per Cluster/Tile	4	4
number of Clusters/number Tiles per Slot	2	6
Subcarriers \times Symbols per Slot	56	72
Data Subcarriers \times Symbols per Slot	48	48
Data Subcarriers \times Symbols per DL and UL Subframe	23,520	12,600
Number of Slots	420	175

TABLE 5: Vocoder parameters [10].

Vocoder	AMR	G.729A	G.711	G.723.1	
				A	B
Source bit rate (kbps)	4.5 to 12.2	8	64	5.3	6.3
Frame duration (millisecond)	20	10	10	30	30
Payload (bytes) (Active, Inactive)	(33, 7)	(20, 0)	(20, 0)	(20, 0)	(20, 0)

5. Overhead Analysis

In this section, we consider both upper and lower layer overheads in detail.

5.1. Upper Layer Overhead. Table 7 which lists the characteristics of our Mobile TV, VoIP, and data workloads includes the type of transport layer used: either Real Time Transport Protocol (RTP) or TCP. This affects the upper layer protocol overhead. RTP over UDP over IP (12 + 8 + 20) or TCP over IP (20 + 20), can result in a per packet header overhead of 40 bytes. This is significant and can severely reduce the capacity of any wireless system.

There are two ways to reduce upper layer overheads and to improve the number of supported users. These are Payload Header Suppression (PHS) and Robust Header Compression (ROHC). PHS is a Mobile WiMAX feature. It allows the sender not to send fixed portions of the headers and can reduce the 40-byte header overhead down to 3 bytes. ROHC, specified by the Internet Engineering Task Force (IETF), is another higher layer compression scheme. It can reduce the higher layer overhead to 1 to 3 bytes. In our analysis, we used ROHC-RTP packet type 0 with R-0 mode. In this mode, all RTP sequence numbers functions are known to the decompressor. This results in a net higher layer overhead of just 1 byte [5, 14, 15].

For small packet size workloads, such as VoIP, header suppression and compression can make a significant impact on the capacity. We have seen several published studies that use uncompressed headers resulting in significantly reduced performance which would not be the case in practice.

PHS or ROHC can significantly improve the capacity and should be used in any capacity planning or estimation.

Note that one option with VoIP traffic is that of silence suppression which if implemented can increase the VoIP capacity by the inverse of fraction of time the user is active (not silent). As a result in this analysis, given a silence suppression option, a number of supported users are twice as much as that without this option.

5.2. Lower Layer Overhead. In this section, we analyze the overheads at MAC and PHY layers. Basically, there is a 6-byte MAC header and optionally several 2-byte subheaders. The PHY overhead can be divided into DL overhead and UL overhead. Each of these three overheads is discussed next.

5.2.1. MAC Overhead. At MAC layer, the smallest unit is MAC protocol data unit (MPDU). As shown in Figure 4, each PDU has at least 6-bytes of MAC header and a variable length payload consisting of a number of optional subheaders, data, and an optional 4-byte Cyclic Redundancy Check (CRC). The optional subheaders include fragmentation, packing, mesh, and general subheaders. Each of these is 2 bytes long.

In addition to generic MAC PDUs, there are bandwidth request PDUs. These are 6 bytes in length. Bandwidth requests can also be piggybacked on data PDUs as a 2-byte subheader. Note that in this analysis, we do not consider the effect of polling and/or other bandwidth request mechanisms.

TABLE 6: Mobile TV workload parameters [12].

Applications	Format	Data rate	Notes
Mobile phone video	H.264 ASP	176 kbps	176×144 , 20 frame per second
Smartphone video	H.264 ASP	324 kbps	320×240 , 24 frame per second
IPTV video	H.264 Baseline	850 kbps	480×480 , 30 frame per second
Sample video trace [11]	MPEG2	350 kbps	Average Packet Size = 984 bytes

TABLE 7: Web workload characteristics.

Parameters	Values
Main page size (bytes)	10,710
Embedded object size (bytes)	7,758
Number of embedded objects	5.64
Reading time (second)	30
Parsing time (second)	0.13
Request size (bytes)	350
Big packet size (bytes)	1,422
Small packet size (bytes)	498
% of big packets	76
% of small packets	24

UL preamble	MAC/BW- REQ header	Other subheader	Data	CRC (optional)
----------------	-----------------------	--------------------	------	-------------------

FIGURE 4: UL burst preamble and MAC PDU (MPDU).

Consider fragmentation and packing subheaders. As shown in Table 9, the user bytes per frame in downlink are 219, 3.5, and 9.1 bytes for Mobile TV, VoIP, and Web, respectively. In each frame, a 2-byte fragmentation subheader is needed for all types of traffic. Packing is not used for the simple scheduler used here.

However, in the enhanced scheduler, given a variation of deadline, packing multiple SDU is possible. Table 9 also shows an example when deadline is put into consideration. In this analysis, the deadlines of Mobile TV, VoIP, and Web traffic are set to 10, 60, and 250 millisecond. As a result, 437.9, 42.0, and 454.9 bytes are allocated per user. These configuration results in one 2-byte fragmentation overhead for Mobile TV and Web traffic but two 2-byte packing overheads with no fragmentation for VoIP. Table 9 also shows the detailed explanation of fragmentation and packing overheads in downlink. Note that the calculation for uplink is very similar.

5.2.2. Downlink Overhead. In DL subframe, the overhead consists of preamble, FCH, DL-MAP, and UL-MAP. The MAP entries can result in a significant amount of overhead since they are repeated 4 times. WiMAX Forum recommends using compressed MAP [3], which reduces the DL-MAP entry overhead to 11 bytes including 4 bytes for CRC [1]. The fixed UL-MAP is 6 bytes long with an optional 4-byte CRC. With a repetition code of 4 and QPSK, both fixed DL-MAP and UL-MAP take up 16 slots.

The variable part of DL-MAP consists of one entry per bursts and requires 60 bits per entry. Similarly, the variable part of UL-MAP consists of one entry per bursts and requires 52 bits per entry. These are all repeated 4 times and use only QPSK MCS. It should be pointed out that the repetition consists of repeating slots (and not bytes). Thus, both DL and UL MAPs entries also take up 16 slots each per burst.

Equation (2) show the details of UL and DL MAPs overhead computation:

$$\begin{aligned}
 UL_MAP(\text{bytes}) &= \frac{48 + 52 \times \#UL_users}{8}, \\
 DL_MAP(\text{bytes}) &= \frac{88 + 60 \times \#DL_users}{8}, \\
 DL_MAP(\text{slots}) &= \left\lceil \frac{UL_MAP}{S_i} \right\rceil \times r, \\
 UL_MAP(\text{slots}) &= \left\lceil \frac{DL_MAP}{S_i} \right\rceil \times r.
 \end{aligned} \tag{2}$$

Here, r is the repetition factor and S_i is the slot size (bytes) given i th modulation and coding scheme. Note that basically QPSK1/2 is used for the computation of UL and DL MAPs.

5.2.3. Uplink Overhead. The UL subframe also has fixed and variable parts (see Figure 1). Ranging and contention are in the fixed portion. Their size is defined by the network administrator. These regions are allocated not in units of slots but in units of *transmission opportunities*. For example, in CDMA initial ranging, one opportunity is 6 subchannels and 2 symbol times.

The other fixed portion is Channel Quality Indication (CQI) and ACKnowledgements (ACKs). These regions are also defined by the network administrator. Obviously, more fixed portions are allocated; less number of slots is available for the user workloads. In our analysis, we allocated three OFDMA symbol columns for all fixed regions.

Each UL burst begins with a UL preamble. Typically, one OFDMA symbol is used for short preamble and two for long preamble. In this analysis, we do not consider one short symbol (a fraction of one slot); however, users can add an appropriate size of this symbol to the analysis.

6. Pitfalls

Many Mobile WiMAX analyses ignore the overheads described in Section 5, namely, UL-MAP, DL-MAP, and MAC overheads. In this section, we show that these overheads have a significant impact on the number of users

TABLE 8: Summary of workload characteristics.

Parameters	Mobile TV	VoIP	Data (Web)
Types of transport layer	RTP	RTP	TCP
Average packet size (bytes)	983.5	20.0	1,200.2
Average data rate (kbps) w/o headers	350.0	5.3	14.5
UL : DL traffic ratio	0	1	0.006
Silence suppression (VoIP only)	N/A	Yes	N/A
Fraction of time user is active		0.5	
ROHC packet type	1	1	TCP
Overhead with ROHC (bytes)	1	1	8
Payload Header Suppression (PHS)	No	No	No
MAC SDU size with header	984.5	21.0	1,208.2
Data rate (kbps) after headers	350.4	5.6	14.6
Bytes/frame per user (DL)	219.0	3.5	9.1
Bytes/frame per user (UL)	0.0	3.5	0.1

TABLE 9: Fragmentation and packing subheaders.

Parameters	Mobile TV	VoIP	Data (Web)
Average packet size with higher level header (bytes)	984.5	21.0	1,208.2
Simple scheduler			
Bytes/5 millisecond frame per user	219.0	3.5	9.1
Number of fragmentation subheaders	1	1	1
Number of packing subheaders	0	0	0
Enhanced scheduler			
Deadline (millisecond)	10	60	250
Bytes/5 millisecond frame per user	437.9	42.0	454.9
Number of fragmentation subheaders	1	0	1
Number of packing subheaders	0	2	0

TABLE 10: Example of capacity evaluation using a simple scheduler.

Parameters	Mobile TV	VoIP	Data (Web)
MAC SDU size with header (bytes)	984.5	21.0	1,208.2
Data rate (kbps) with upper layer headers	350.4	5.6	14.6
(a) DL			
Bytes/5 millisecond frame per user (DL)	219.0	3.5	9.1
Number of fragmentation subheaders	1	1	1
Number of packing subheaders	0	0	0
DL data slots per user with MAC header + packing and fragmentation subheaders	38	2	3
Total slots per user (Data + DL-MAP IE + UL-MAP IE)	46	18	19
Number of users (DL)	9	35	33
(b) UL			
Bytes/5 millisecond frame per user (UL)	0.0	3.5	0.1
number of fragmentation subheaders	0	1	1
number of packing subheaders	0	0	0
UL data slots per user with MAC header + packing and fragmentation subheaders	0	2	2
Number of users (UL)	8	87	87
Number of users (min of UL and DL)	9	35	33
Number of users with silence suppression	9	70	33

supported. Since some of these overheads depend upon the number of users, the scheduler needs to be aware of this additional need while admitting and scheduling the users [4, 17]. We present two case studies. The first one assumes an error-free channel while the second extends the results to a case in which different users have different error rates due to channel conditions.

6.1. Case Study 1: Error-Free Channel. Given the user workload characteristics and the overheads discussed so far, it is straightforward to compute the system capacity for any given workload. Using the slot capacity indicated in Table 3, for various MCSs, we can compute the number of users supported.

One way to compute the number of users is simply to divide the channel capacity by the bytes required by the user payload and overhead [4]. This is shown in Table 10. The table assumes QPSK 1/2 MCS for all users. This can be repeated for other MCSs. The final results are as shown in Figure 5. The number of users supported varies from 2 to 82 depending upon the workload and the MCS.

The number of users depends upon the available capacity which depends on the MAP overhead, which in turn is determined by the number of users. To avoid this recursion, we use (3) to (5) that give a very good approximation for the number of supported users using a ceiling function:

$$\begin{aligned} \#DL_slots = & \left\lceil \frac{DL_MAP + CRC + \#DL_users \times DIE}{S_i} \right\rceil \times r \\ & + \left\lceil \frac{UL_MAP + CRC + \#UL_users \times UIE}{S_i} \right\rceil \times r \\ & + \#DL_users \times \left\lceil \frac{D}{S_k} \right\rceil, \end{aligned} \quad (3)$$

$$\#UL_slots = \#UL_users \times \left\lceil \frac{D}{S_k} \right\rceil, \quad (4)$$

$$D = B + MAC_{header} + \text{Subheaders}. \quad (5)$$

Here, D is the data size (per frame) including overheads, B is the bytes per frame, and MAC_{header} is 6 bytes. Subheaders are fragmentation and packing subheaders, 2 bytes each if present. DIE and UIE are the sizes of downlink and uplink map information elements (IEs). Note that DL_MAP and UL_MAP are fixed MAP parts and also in terms of bytes. Again, r is the repetition factor and S_i is the slot size (bytes) given i th modulation and coding scheme. number DL_slots is the total number of DL slots without preamble and number UL_slots are the total number of UL slots without ranging, ACK, and CQICH.

For example, consider VoIP with QPSK 1/2 (slot size = 6 bytes) and repetition of four. Equation (3) results 35 users in the downlink. The derivation is as follows:

$$\begin{aligned} \#DL_slots &= 420 \\ &= \left\lceil \frac{11 + 4 + \#DL_users \times 60/8}{6} \right\rceil \times 4 \\ &\quad + \left\lceil \frac{6 + 4 + \#UL_users \times 52/8}{6} \right\rceil \\ &\quad \times 4 + \#DL_users \times \left\lceil \frac{11.5}{6} \right\rceil. \end{aligned} \quad (6)$$

For uplink, from (4) and (5), the number of UL users is 87:

$$\#UL_slots = 175 = \#UL_users \times \left\lceil \frac{3.5 + 6 + 2}{6} \right\rceil. \quad (7)$$

Finally, after calculating the number of supported users for both DL and UL, the total number of supported users is the minimum of those two numbers. In this example, the total number of supported users is 35, (minimum of 35 and 87). In this case, the downlink is the bottleneck mostly due to the large overhead. Together with silence suppression, the absolute number of supported users can be up to $2 \times 35 = 70$ users. Figure 5 shows the number of supported users for various MCSs.

The main problem with the analysis presented above is that it assumes that every user is scheduled in every frame. Since there is a significant per burst overhead, this type of allocation will result in too much overhead and too little capacity. Also, since every packet (SDU) is fragmented, a 2-byte fragmentation subheader is added to each MAC PDU.

What we discussed above is a common pitfall. The analysis assumes a dumb scheduler. A smarter scheduler will try to aggregate payloads for each user and thus minimizing the number of bursts. We call this the enhanced scheduler. It works as follows. Given n users with any particular workload, we divide the users in k groups of n/k users each. The first group is scheduled in the first frame; the second group is scheduled in the second frame, and so on. The cycle is repeated every k frames. Of course, k should be selected to match the delay requirements of the workload.

For example, with VoIP users, a VoIP packet is generated every 30 millisecond, but assuming 60 millisecond is an acceptable delay, we can schedule a VoIP user every 12th Mobile WiMAX frame (recall that each Mobile WiMAX frame is 5 millisecond) and send two VoIP packets in one frame as compared to the previous scheduler which would send 1/6th of the VoIP packet in every frame and thereby aggravating the problem of small payloads. Two 2-byte packing headers have to be added in the MAC payload along with the two SDUs.

Table 11 shows the capacity analysis for the three workloads with QPSK 1/2 MCS and the enhanced scheduler. The results for other MCSs can be similarly computed. These results are plotted in Figure 6. Note that the number of users supported has gone up significantly. Compared to Figure 5, there is a capacity improvement by a factor of 1 to 20 depending upon the workload and the MCS.

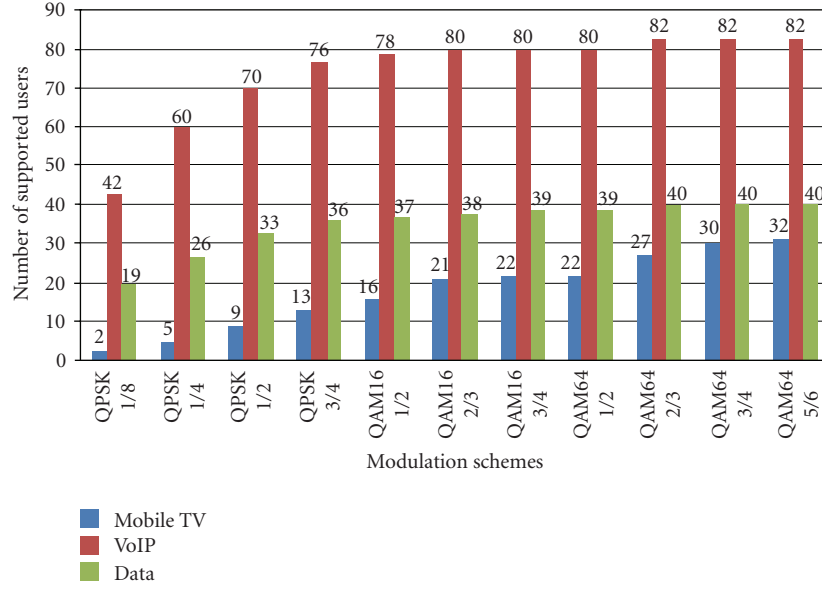


FIGURE 5: Number of users supported in a lossless channel (Simple scheduler).

TABLE 11: Example of capacity evaluation using an enhanced scheduler.

Parameters	Mobile TV	VoIP	Data (Web)
MAC SDU size with header (bytes)	984.5	21.0	1,208.2
Data rate (kbps) with upper layer headers	350.4	2.8	14.6
Deadline (millisecond)	10	60	250
(a) DL			
Bytes/5 millisecond frame per user (DL)	437.9	42.0	454.9
Number of fragmentation subheaders	1	0	1
Number of packing subheaders	0	2	0
DL data slots per user with MAC header + packing and fragmentation subheaders	75	9	78
Total slots per user (Data + DL-MAP IE + UL-MAP IE)	83	25	94
Number of users (DL)	10	269	233
(b) UL			
Bytes/5 millisecond frame per user (UL)	0.0	42.0	2.9
Number of fragmentation subheaders	1	0	1
Number of packing subheaders	0	2	0
UL data slots per user with MAC header + packing and fragmentation subheaders	0	9	2
Number of users (UL)	8	228	4350
Net number of users (min of UL and DL)	10	228	233
Number of users with silence suppression	10	456	233

Proper scheduling can change the capacity by an order of magnitude. Making less frequent but bigger allocations can reduce the overhead significantly.

The number of supported users for this scheduler is derived from the same equations that were used with the simple scheduler. However, the enhanced scheduler allocates as large size as possible given the deadlines. For example, for Mobile TV with a 10-millisecond deadline, instead of

219 bytes, the scheduler allocates 437.9 bytes within a single frame and for VoIP with 60-millisecond deadline, instead of 3.5 bytes per frame, it allocates 42 bytes and that results in 2 packing overheads instead of 1 fragmentation overhead.

In Table 11, the number of supported users for VoIP is 228. This number is based on the fact that 42 bytes are allocated for each user every 60 millisecond:

$$\left\lceil \frac{\text{\#slots_subframe}}{\text{\#slots_aggregated_users}} \right\rceil \times \frac{\text{deadline}}{5 \text{ millisecond}} \quad (8)$$

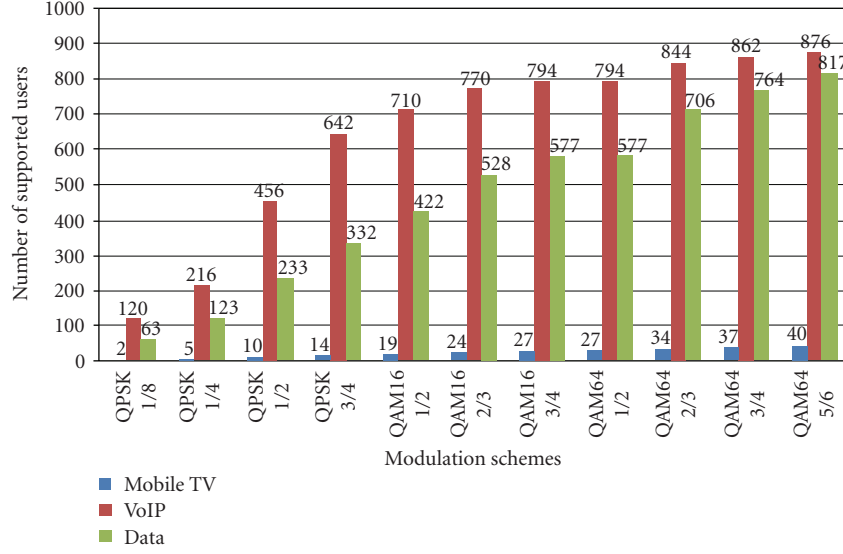


FIGURE 6: Number of users supported in a lossless channel (Enhanced Scheduler).

TABLE 12: Simulation parameters [16].

Parameters	Value
Channel model	ITU Veh-B (6 taps) 120 km/hr
Channel bandwidth	10 MHz
Frequency band	2.35 GHz
Forward Error Correction	Convolution Turbo Coding
Bit Error Rate threshold	10^{-5}
MS receiver noise figure	6.5 dB
BS antenna transmit power	35 dBm
BS receiver noise figure	4.5 dB
Path loss PL(distance)	$37 \times \log_{10}(\text{distance}) + 20 \times \log_{10}(\text{frequency}) + 43.58$
Shadowing	Log normal with $\sigma = 10$
number of sectors per cell	3
Frequency reuse	1/3

TABLE 13: Percent MCS for 1×1 and 2×2 antennas [16].

Average MCS	1 Antenna		2 Antenna	
	%DL	%UL	%DL	%UL
FADE	4.75	1.92	3.03	1.21
QPSK 1/8	7.06	3.54	4.06	1.68
QPSK 1/4	16.34	12.46	14.64	8.65
QPSK 1/2	15.30	20.01	13.15	14.05
QPSK 3/4	12.14	21.23	10.28	15.3
QAM-16 1/2	20.99	34.33	16.12	29.97
QAM-16 2/3	0.00	0.00	0.00	0.00
QAM-16 3/4	9.31	5.91	14.18	22.86
QAM-64 1/2	0.00	0.00	0.00	0.00
QAM-64 2/3	14.11	0.59	24.53	6.27

With the configuration in Table 11, the number of supported users is $\lceil 175/9 \rceil \times 60/5 = 228$ users. With silence

suppression, the absolute number of supported users is $2 \times 228 = 456$. Note that the number of DL users is computed using (3), (4), and (5), and then (9) can be applied. The calculations for Mobile TV and Data are similar to that for VoIP.

The per-user overheads impact the downlink capacity more than the uplink capacity. The downlink subframe has DL-MAP and UL-MAP entries for all DL and UL bursts and these entries can take up a significant part of the capacity and so minimizing the number of bursts increases the capacity.

Note that there is a limit to aggregation of payloads and minimization of bursts. First, the delay requirements for the payload should be met and so a burst may have to be scheduled even if the payload size is small. In these cases, multiuser bursts in which the payload for multiple users is aggregated in one DL burst with the same MCS can help reduce the number of bursts. This is allowed by the IEEE 802.16e standards and applies only to the downlink bursts.

TABLE 14: Number of supported users in a lossy channel.

Workload	1 Antenna		2 Antenna	
	Simple scheduler	Enhanced scheduler	Simple scheduler	Enhanced scheduler
Mobile TV	14	16	17	20
VoIP	76	672	78	720
Data	36	369	37	438

The second consideration is that the payload cannot be aggregated beyond the frame size. For example, with QPSK 1/2, a Mobile TV application will generate enough load to fill the entire DL subframe every 10 millisecond or every 2 frames. This is much smaller than the required delay of 30 millisecond between the frames.

6.2. Case Study 2: Imperfect Channel. In Section 6.1, we saw that the aggregation has more impact on performance with higher MCSs (which allow higher capacity and hence more aggregation). However, it is not always possible to use these higher MCSs. The MCS is limited by the quality of the channel. As a result, we present a capacity analysis assuming a mix of channels with varying quality resulting in different levels of MCS for different users.

Table 12 lists the channel parameters used in a simulation by Leiba et al. [16]. They showed that under these conditions, the number of users in a cell which were able to achieve any particular MCS was as listed in Table 13. Two cases are listed: single antenna systems and two antenna systems.

Average bytes per slot in each direction can be calculated by summing the product (percentage users with an MCS \times number of bytes per slot for that MCS). For 1 antenna systems this gives 10.19 bytes for the downlink and 8.86 bytes for the uplink. For 2 antenna systems, we get 12.59 bytes for the downlink and 11.73 bytes for the uplink.

Table 14 shows the number of users supported for both simple and enhanced schedulers. The results show that the enhanced scheduler still increases the number of users by an order of magnitude, especially for VoIP and data users.

7. Conclusions

In this paper, we explained how to compute the capacity of a Mobile WiMAX system and account for various overheads. We illustrated the methodology using three sample workloads consisting of Mobile TV, VoIP, and data users. Analysis such as the one presented in this paper can be easily programmed in a simple program or a spread sheet and effect of various parameters can be analyzed instantaneously. This can be used to study the sensitivity to various parameters so that parameters that have significant impact can be analyzed in detail by simulation. This analysis can also be used to validate simulations.

However, there are a few assumptions in the analysis such as the effect of bandwidth request mechanism, two-dimensional downlink mapping, and the imprecise calculation of slot-based versus bytes-based. Moreover, we do

not consider (H)ARQ [18]. In addition, the number of supported users is calculated with the assumption that there is only one traffic type. Finally, fixed UL-MAP is always in the DL subframe though there is no UL traffic such as Mobile TV [4].

We showed that proper accounting of overheads is important in capacity estimation. A number of methods are available to reduce these overheads and these should be used in all deployments. In particular, robust header compression or payload header suppression and compressed MAPs are examples of methods for reducing the overhead.

Proper scheduling of user payloads can change the capacity by an order of magnitude. The users should be scheduled so that their numbers of bursts are minimized while still meeting their delay constraint. This reduces the overhead significantly particularly for small packet traffic such as VoIP.

We also showed that our analysis can be used for loss-free channel as well as for noisy channels with loss.

Acknowledgment

This work was sponsored in part by a grant from Application Working Group of WiMAX Forum. “WiMAX,” “Mobile WiMAX,” “Fixed WiMAX,” “WiMAX Forum,” “WiMAX Certified,” “WiMAX Forum Certified,” the WiMAX Forum logo and the WiMAX Forum Certified logo are trademarks of the WiMAX Forum.

References

- [1] IEEE P802.16Rev2/D2, “DRAFT Standard for Local and metropolitan area networks,” Part 16: Air Interface for Broadband Wireless Access Systems, p. 2094, December 2007.
- [2] C. So-In, R. Jain, and A.-K. Tamimi, “AWG Analytical Model for Application Capacity Planning over WiMAX V0.8,” WiMAX Forum, Application Working Group (AWG) Contribution, September 2009, <http://cse.wustl.edu/~jain/papers/capmodel.xls>.
- [3] WiMAX Forum, “WiMAX System Evaluation Methodology V2.1,” p. 230, July 2008, <http://www.wimaxforum.org/resources/documents/technical>.
- [4] C. So-In, R. Jain, and A.-K. Tamimi, “Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey,” *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 156–171, 2009.
- [5] C. Eklund, R.-B. Marks, S. Ponnuswamy, K.-L. Stanwood, and N.-V. Waes, *WirelessMAN Inside the IEEE 802.16 Standard for Wireless Metropolitan Networks*, IEEE Standards Information Network/IEEE Press, Piscataway, NJ, USA, 2006.

- [6] G. Jeffrey, J. Andrews, A. Arunabha-Ghosh, and R. Muhamed, *Fundamentals of WiMAX Understanding Broadband Wireless Networking*, Prentice-Hall PTR, Upper Saddle River, NJ, USA, 2007.
- [7] L. Nuaymi, *WiMAX: Technology for Broadband Wireless Access*, John Wiley & Sons, New York, NY, USA, 2007.
- [8] H. Yaghoobi, "Scalable OFDMA physical layer in IEEE 802.16 wirelessMAN," *Intel Technology Journal*, vol. 8, no. 3, pp. 202–212, 2004.
- [9] R. Jain, C. So-In, and A.-K. Tamimi, "System-level modeling of IEEE 802.16E mobile WiMAX networks: key issues," *IEEE Wireless Communications*, vol. 15, no. 5, pp. 73–79, 2008.
- [10] R. Srinivasan, T. Papathanassiou, and S. Timiri, "Mobile WiMAX VoIP capacity system level simulations," Application Working Group, WiMAX Forum, Beaverton, Ore, USA, 2007.
- [11] A.-K. Tamimi, R. Jain, and C. So-In, "SAM: simplified seasonal ARIMA model for wireless broadband access enabled mobile devices," in *Proceedings of IEEE International Symposium on Multimedia (ISM '08)*, pp. 178–183, Berkeley, Calif, USA, December 2008.
- [12] D. Ozdemir and F. Retnasothie, "WiMAX capacity estimation for triple play services including mobile TV, VoIP and Internet," Application Working Group, WiMAX Forum, Beaverton, Ore, USA, 2007.
- [13] 3rd Generation Partnership Project, "HTTP and FTP traffic model for 1xEV-DV simulations," *3GPP2-C50-EVAL-2001022-0xx*, 2001.
- [14] G. Pelletier, K. Sandlund, L.-E. Jonsson, and M. West, "RObust Header Compression (ROHC): A Profile for TCP/IP (ROHC-TCP)," *RFC 4996*, January 2006.
- [15] L.-E. Jonsson, G. Pelletier, and K. Sandlund, "Framework and four profiles: RTP, UDP, ESP and uncompressed," *RFC 3095*, July 2001.
- [16] Y. Leiba, Y. Segal, Z. Hadad, and I. Kitroser, "Coverage/capacity simulations for OFDMA PHY in with ITU-T channel model," type C802.16d-03/78, IEEE, November 2004.
- [17] C. So-In, R. Jain, and A.-K. Tamimi, "A deficit round robin with fragmentation scheduler for IEEE 802.16e mobile WiMAX," in *Proceedings of IEEE Sarnoff Symposium (SARNOFF '09)*, pp. 1–7, Princeton, NJ, USA, March-April 2009.
- [18] A. Sayenko, O. Alanen, and T. Hamalainen, "ARQ aware scheduling for the IEEE 802.16 base station," in *Proceedings of IEEE International Conference on Communications (ICC '08)*, pp. 2667–2673, Beijing, China, May 2008.