# Capacity Evaluation for IEEE 802.16e Mobile WiMAX

Chakchai So-In, *Student Member, IEEE,* Raj Jain, *Fellow, IEEE,* and Abdel-Karim Tamimi, *Student Member, IEEE*

***Abstract*—We present a simple analytical method for capacity evaluation of IEEE 802.16e Mobile WiMAX$^{TM}$ 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.**

***Index Terms*—WiMAX, IEEE 802.16, Mobile WiMAX, IEEE 802.16e, Capacity Planning, Capacity Evaluation, Application Performance, Overhead, Mobile TV, VoIP**

## I. Introduction

IEEE 802.16e Mobile WiMAX$^{TM}$ 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 II, we present an overview of Mobile WiMAX physical layer (PHY). Understanding this is important for performance modeling. In Section III, 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 IV. Our analysis is general and can be used for any other application workload. Section V explains both upper and lower layer overheads and ways to reduce those overheads. The number of users supported for the three workloads are finally presented in Section VI. 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 VII.

## II. 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], [4], [5], [6], [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).

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.

C. So-In, R. Jain, and A. Tamimi is with the Department of Computer Science and Engineering, Washington Univeristy in St.Louis, St.Louis, MO, 63130. E-mail: cs5, jain, and aa7@cse.wustl.edu

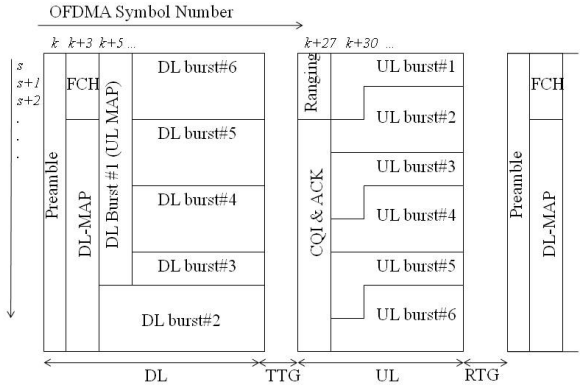



Fig. 1. A Sample OFDMA Frame Structure

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 Fig. 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 in to 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). Figs. 2 and 3 show slot formation for PUSC. In uplink (Fig. 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

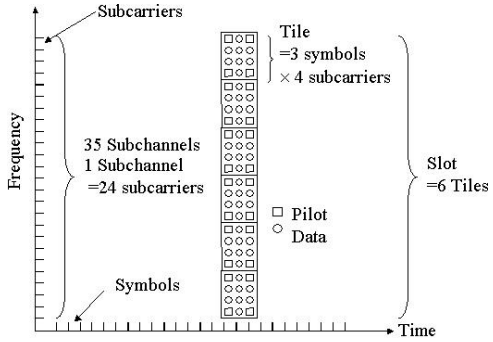



Fig. 2. Symbols, Tiles, and Slots in Uplink PUSC

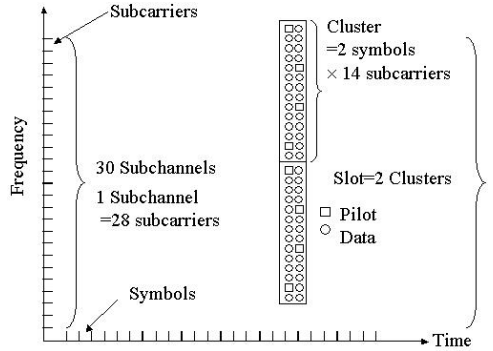



Fig. 3. Symbols, Clusters, and Slots in Downlink PUSC

form a subchannel. Therefore, at 10 MHz, 1024 subcarriers form 35 UL subchannels. The slot formation in downlink is different and is shown in Fig. 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.

The Mobile WiMAX DL subframe, as shown in Fig. 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.

## III. Mobile WiMAX Configuration Parameters and Characteristics

The key parameters of Mobile WiMAX PHY are summarized in Table I through III.

Table I 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, $10.93kHz \times 1024 = 10MHz \times \frac{28}{25}$. This table shows that at 10 MHz

the OFDMA symbol time is 102.8 $\mu$s and so there are 48.6 symbols in a 5 ms 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 consists of a total of $14\times30$ or 420 slots. The UL subframe will consist of $6\times35$ or 210 slots.

Table II 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 III lists the number of bytes per slot for various MCS values. For each MCS, the number of bytes is equal to [#bits per symbols $\times$ Coding Rate $\times$ 48 data subcarriers and symbols per slot / 8 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 ms 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 IV.

## IV. 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 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 ms. Note that other vocoder parameters can be also used and they are listed in Table V.

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 ms 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 VI 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 VII.

The characteristics of the three workloads are summarized in Table VIII. 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) are presented in the next section. Given ROHC, the data rate with higher level headers ($R_{withHeader}$) is calculated by:

$$R_{withHeader} = R \times \frac{(MSDU + Header)}{MSDU} \tag{1}$$

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 $R_{withHeader} \times frame\_duration$. 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.

## V. Overhead Analysis

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

### A. Upper Layer Overhead

Table VII 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 $(20+20)$, both 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 to not 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.–*

TABLE I
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 ($\Delta f$, kHz) | 10.93 | | | | 7.81 | | 9.76 |
| Useful symbol time ($T_b = \frac{1}{\Delta f}$, $\mu s$) | 91.4 | | | | 128 | | 102.4 |
| Guard time ($T_g = \frac{T_b}{8}$, $\mu s$) | 11.4 | | | | 16 | | 12.8 |
| OFDMA symbol time ($T_s = T_b + T_g$, $\mu s$) | 102.8 | | | | 144 | | 115.2 |

TABLE II
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 |
| # of guard subcarriers | 43 | N/A | 91 | 183 | 367 |
| # of used subcarriers | 85 | N/A | 421 | 841 | 1,681 |
| # of pilot subcarriers | 12 | N/A | 60 | 120 | 240 |
| # 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 |
| # of guard subcarriers | 31 | N/A | 103 | 183 | 367 |
| # of used subcarriers | 97 | N/A | 409 | 841 | 1,681 |

TABLE III
SLOT CAPACITY FOR VARIOUS MCSS

| **MCS** | **Bits per symbol** | **Coding Rate** | **DL bytesper slot** | **UL bytesper 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 |

TABLE IV
MOBILE WIMAX SYSTEM CONFIGURATIONS

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

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.

### *B. 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

TABLE V
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 (ms) | 20 | 10 | 10 | 30 | 30 |
| Payload (bytes) (Active, Inactive) | (33, 7) | (20, 0) | (20, 0) | (20, 0) | (20, 0) |

TABLE VI
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 VII
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 |

TABLE VIII
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 |

three overheads is discussed next.

*1) MAC Overhead:* At MAC layer, the smallest unit is MAC protocol data unit (MPDU). As shown in Fig. 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.

Consider fragmentation and packing subheaders. As shown in Table IX, 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.

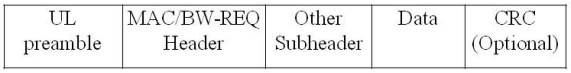



Fig. 4. UL burst preamble and MAC PDU (MPDU)

However, in the enhanced scheduler, given a variation of deadline, packing multiple SDU is possible. Table IX also shows an example when deadline is put into consideration. In this analysis, the deadline of Mobile TV, VoIP, and Web traffic are set to 10, 60, and 250 ms. 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 IX also shows the detailed explanation of

fragmentation and packing overheads in downlink. Note that the calculation for uplink is very similar.

*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.

Equations (2) to (5) show the details of UL and DL MAPs overhead computation.

$$UL\_MAP(bytes) = \frac{48 + 52 \times \#UL\_users}{8} \quad (2)$$

$$DL\_MAP(bytes) = \frac{88 + 60 \times \#DL\_users}{8} \quad (3)$$

$$DL\_MAP(slots) = \left\lceil \frac{UL\_MAP}{S_i} \right\rceil \times r \quad (4)$$

$$UL\_MAP(slots) = \left\lceil \frac{DL\_MAP}{S_i} \right\rceil \times r \quad (5)$$

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

*3) Uplink Overhead:* The UL subframe also has fixed and variable parts (See Fig. 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 (ACK). 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.

## VI. Pitfalls

Many Mobile WiMAX analyses ignore the overheads described in Section V, 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 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], [16]. 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.

### A. 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 III, 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 X. The table assumes QPSK 1/2 MCS for all users. This can be repeated for other MCSs. The final results are as shown in Fig. 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 equations (6) to (8) that give a very good approximation for the number of supported users using a ceiling function:

$$\begin{aligned} \#DL\_slots = \\ \left\lceil \frac{DL\_MAP + CRC + \#DL\_users \times DIE}{S_i} \right\rceil \times r \\ + \left\lceil \frac{UL\_MAP + CRC + \#UL\_users \times UIE}{S_i} \right\rceil \times r + \\ \#DL\_users \times \left\lceil \frac{D}{S_k} \right\rceil \end{aligned} \quad (6)$$

$$\#UL\_slots = \#UL\_users \times \lceil \frac{D}{S_k} \rceil \quad (7)$$

$$D = B + MAC_{header} + Subheaders \quad (8)$$

Here, $D$ is the data size (per frame) including overheads, $B$ is the bytes per frame, $MAC_{header}$ 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 $ith$ modulation and coding scheme. $\#DL\_slots$ is the total number of DL slots without preamble and $\#UL\_slots$ 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 (6) results 35 users in the downlink. The derivation is as follows:

$$\#DL\_slots = 420 = \left\lceil \frac{11 + 4 + \#DL\_users \times \frac{60}{8}}{6} \right\rceil \times 4 + \left\lceil \frac{6 + 4 + \#UL\_users \times \frac{52}{8}}{6} \right\rceil \times 4 + \#DL\_users \times \left\lceil \frac{11.5}{6} \right\rceil$$

TABLE IX
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 ms 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 (ms) | 10 | 60 | 250 |
| Bytes/5 ms frame per user | 437.9 | 42.0 | 454.9 |
| Number of fragmentation subheaders | 1 | 0 | 1 |
| Number of packing subheaders | 0 | 2 | 0 |

For uplink, from equations 7 and 8, the number of UL users is 87.

$$\#UL\_slots = 175 = \#UL\_users \times \left\lceil \frac{3.5+6+2}{6} \right\rceil$$

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. Fig. 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 $\frac{n}{k}$ 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 ms but assuming 60 ms is an acceptable delay, we can schedule a VoIP user every $12th$ Mobile WiMAX frame (recall that each Mobile WiMAX frame is 5 ms) 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 XI 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 Fig. 6. Note that the number of users supported has gone up significantly. Compared to Fig. 5, there is a capacity improvement by a factor of 1 to 20 depending upon the workload and the MCS.

*–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-ms deadline, instead of 219 bytes, the scheduler allocates 437.9 bytes within a single frame and for VoIP with 60 ms 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 XI, 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 ms:

$$\left\lceil \frac{\#slots\_subframe}{\#slots\_aggregated\_users} \right\rceil \times \frac{deadline}{5ms} \tag{9}$$

With the configuration in Table XI, the number of supported users is $\lceil \frac{175}{9} \rceil \times \frac{60}{5} = 228$ 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 equations 6, 7, and 8; and then equation 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, multi-user 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.

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 ms or every 2 frames. This is much smaller than the required delay of 30 ms between the frames.

TABLE X
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 ms 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/5ms frame per user (UL) | 0.0 | 3.5 | 0.1 |
| # of fragmentation subheaders | 0 | 1 | 1 |
| # 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** |

TABLE XI
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 (ms)** | 10 | 60 | 250 |
| (a) DL | | | |
| Bytes/5 ms 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 ms 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** |

### *B. Case Study 2: Imperfect Channel*

In Section A, 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 XII lists the channel parameters used in a simulation by Leiba et al. [17]. 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 XIII. 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 XIV 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.

## VII. 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

TABLE XII
SIMULATION PARAMETERS [17]

| **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) | $37 \times log10(distance) + 20 \times log10(frequency) + 43.58$ |
| Shadowing | Log normal with $\sigma$ =10 |
| # of sectors per cell | 3 |
| Frequency reuse | 1/3 |

TABLE XIII
PERCENT MCS FOR $1 \times 1$ AND $2 \times 2$ ANTENNAS [17]

| **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 |

TABLE XIV
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 |

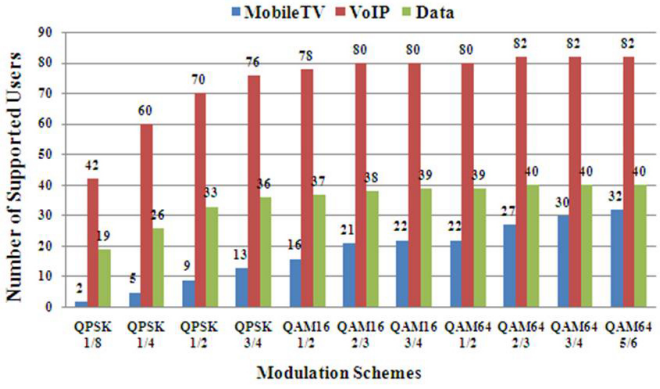



Fig. 5. Number of users supported in a lossless channel (Simple scheduler)

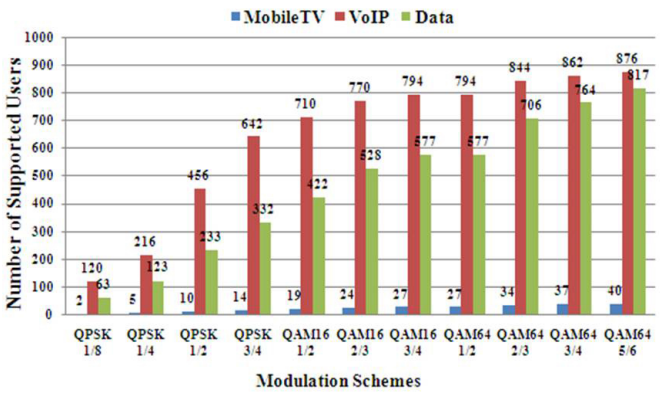



Fig. 6. Number of users supported in a lossless channel (Enhanced Scheduler)

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 vs. 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.

## REFERENCES


[1] IEEE P802.16Rev2/D2, "DRAFT Standard for Local and metropolitan area networks," Part 16: Air Interface for Broadband Wireless Access Systems, Dec. 2007, 2094 pp.

[2] C. So-In, R. Jain, and A. Al-Tamimi, "AWG Analytical Model for Application Capacity Planning over WiMAX V0.8," *WiMAX Forum, Application Working Group (AWG) Contribution*, Sept. 2009. Available: http://www.cse.wustl.edu/jain/papers/capmodel.xls

[3] WiMAX Forum, "WiMAX System Evaluation Methodology V2.1," July 2008, 230 pp. Available: http://www.wimaxforum.org/resources/documents/technical

[4] C. So-In, R. Jain, and A. Al-Tamimi, "Scheduling in IEEE 802.16e WiMAX Networks: Key Issues and a Survey," *IEEE J. on Selected Areas in Comm. (Special Issue on Broadband Access Networks: Architectures and Protocols)*, vol. 27, no. 2, pp. 156-171, Feb. 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*, May 2006, 400 pp.

[6] G. Jeffrey, J. Andrews, A. Arunabha-Ghosh, and R. Muhamed, "Fundamentals of WiMAX Understanding Broadband Wireless Networking," *Prentice Hall PTR*, Mar. 2007, 496 pp.

[7] L. Nuaymi, "WiMAX: Technology for Broadband Wireless Access," *Wiley*, Mar. 2007, 310 pp.

[8] H. Yaghoobi, "Scalable OFDMA Physical Layer in IEEE 802.16 WirelessMAN," *Intel Technology J.*, vol. 8, no. 3, pp. 202-212, Aug. 2004.

[9] R. Jain, C. So-In, and A. Al-Tamimi, "System Level Modeling of IEEE 802.16e Mobile WiMAX Networks: Key Issues," *IEEE Wireless Comm. Mag.*, vol. 8, no. 3, pp. 202-212, Aug. 2004.

[10] R. Srinivasan, T. Papathanassiou, and S. Timiri, "Mobile WiMAX VoIP Capacity System Level Simulations," *WiMAX Forum, Application Working Group (AWG)*, Mar. 2007.

[11] A. Al-Tamimi, R. Jain, and C. So-In, "SAM: Simplified Seasonal ARIMA Model for Wireless Broadband Access Enabled Mobile Devices" in *Proc. IEEE Int. Symp. on Multimedia*, 2008, pp. 178-183.

[12] D. Ozdemir and F. Retnasothie, "WiMAX Capacity Estimation for Triple Play Services including Mobile TV, VoIP, and Internet," *WiMAX Forum, Application Working Group (AWG)*, June 2007.

[13] 3rd Generation Partnership Project, "HTTP and FTP Traffic Model for 1xEV-DV Simulations," *3GPP2-C50-EVAL-2001022-0xx*, 2001.

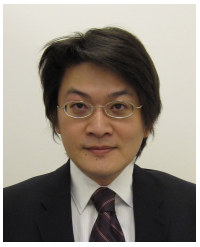
[14] G. Pelletier, K. Sandlund, L-E. Jonsson, and M. West, "RObust Header Compression (ROHC): A Profile for TCP/IP (ROHC-TCP)," *RFC 4996*, Jan. 2006.

[15] L-E. Jonsson, G. Pelletier, and K. Sandlund, "Framework and four profiles: RTP, UDP, ESP and uncompressed," *RFC 3095*, July 2001.

[16] C. So-In, R. Jain, and A. Al-Tamimi, "A Deficit Round Robin with Fragmentation Scheduler for IEEE 802.16e Mobile WiMAX," in *Proc. IEEE Sarnoff Symp.*, 2009, pp. 1-7. Available: http://www.cse.wustl.edu/~jain/papers/drrf.htm

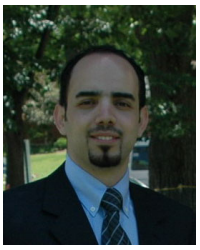
[17] Y. Leiba, Y. Segal, Z. Hadad, and I. Kitroser, "Coverage/ Capacity simulations for OFDMA PHY in with ITU-T channel model," *IEEE C802.16d-03/78*, Nov. 2004, 24 pp.

[18] A. Sayenko, O. Alanen, and T. Hamalainen, "ARQ aware scheduling for the IEEE 802.16 base station," in *Proc. IEEE Computer Communication Conf.*, 2008, pp. 2667-2673.

**Chakchai So-In** Chakchai So-In received his B.Eng. and M.Eng. degrees in Computer Engineering from Kasetsart University, Bangkok, Thailand in 1999 and 2001 respectively. In 2003, He was an internetworking trainee in a CNAP program at Nanyang Technological University, Singapore and obtained Cisco Career Certifications e.g., CCNP and CCDP. He also received M.S. in Computer Engineering from the Department of Computer Science and Engineering, Washington University in St. Louis (WUSTL), MO USA in 2006. In summer 2006, He was a student intern at mobile IP division, Cisco Systems, CA USA. He was also a student intern at WiMAX Forum during summer 2008. Currently he is working toward his Ph.D. degree at WUSTL. His research interests include architectures for Future Wireless Networks/Next Generation Wireless Networks; congestion control in high speed networks; protocols to support network and transport Mobility, Multihoming, and Privacy; and Quality of Service in broadband wireless access networks i.e., Mobile WiMAX and LTE. He is an IEEE student member.

**Raj Jain** Raj Jain is a Professor of Computer Science and Engineering at Washington University in St. Louis. He was one of the Co-founders of Nayna Networks, Inc - a next generation telecommunications systems company in San Jose, CA. Previously, he was a Senior Consulting Engineer at Digital Equipment Corporation in Littleton, Mass and then a professor of Computer and Information Sciences at Ohio State University in Columbus, Ohio. Dr. Jain is a Fellow of IEEE, a Fellow of ACM and ranks among the top 50 in Citeseer's list of Most Cited Authors in Computer Science. He is the author of "Art of Computer Systems Performance Analysis," which won the 1991 "Best-Advanced How-to Book, Systems" award from Computer Press Association. His fourth book entitled " High-Performance TCP/IP: Concepts, Issues, and Solutions," was published by Prentice Hall in November 2003. He is also a winner of ACM SIGCOMM Test of Time award.



**Abdel-Karim Tamimi** Abdel Karim Al Tamimi is a PhD candidate in Computer Engineering at Washington University in St. Louis, MO. Abdel Karim received a BA degree in computer engineering from Yarmouk University in Jordan. His college education was supported by a full scholarship given to excellent students from the High Education Ministry [1999-2004]. During this period, he worked on several projects with different parties tackling problems related to AI, Networking, Computer Security and Digital Imaging. After graduating, he held a full time position as a teaching assistant for several courses in Yarmouk University [2004-2005]. Since then, he has been awarded a full scholarship to peruse his Master and PhD degrees in Computer Engineering at Washington University in St. Louis, where he obtained his master degree [2007].