Tran, M., Halls, DE., Nix, AR., Doufexi, A., & Beach, MA. (2009). Mobile WiMAX: MIMO performance analysis from a Quality of Service (QoS) viewpoint. In IEEE Wireless Communications and Networking Conference (WCNC2009), Budapest, Hungary (pp. 1 - 6). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/WCNC.2009.4917808

Peer reviewed version

Link to published version (if available):
10.1109/WCNC.2009.4917808

Link to publication record in Explore Bristol Research
PDF-document
Mobile WiMAX: MIMO Performance Analysis from a Quality of Service (QoS) Viewpoint

Mai Tran, David Halls, Andrew Nix, Angela Doufexi and Mark Beach
Centre for Communications Research, University of Bristol
Bristol, United Kingdom

Abstract—The end-to-end performance of applications in broadband wireless networks is a key concern from the user perspective. Mobile WiMAX is designed to support a wide range of different applications including media streaming, voice-over-IP (VoIP), video conferencing, and web browsing. All of these applications require different levels of Quality of Service (QoS) and this imposes a variety of different performance requirements on the WiMAX Physical (PHY) and Medium Access Control (MAC) layers. For example, TCP-based applications such as FTP or web-browsing use Automatic-Repeat-reQuest (ARQ) to tolerate a relatively high PHY Packet Error Rate (PER), whilst a UDP-based application such as video streaming requires very low PER due to the lack of ARQ. This means that the same WiMAX system configuration needs to be able to provide different performance levels for different types of applications in order to operate efficiently. This paper analyses and compares the performance of a Multi-Input Multi-Output (MIMO) mobile WiMAX system in terms of achievable throughput and operating range for different PHY PER requirements. The paper shows that TCP applications can achieve higher throughputs and operating ranges due to higher PER tolerance. It also shows that this important QoS parameter affects the AMC switching points. Cross-layer interaction is required between the PHY and the scheduler in the MAC layer in order to efficiently fulfil the various QoS requirements at the application layer.

Index Terms—802.16e, MIMO, BWA, Mobile WiMAX, MAC

I. INTRODUCTION

The first WiMAX systems were based on the IEEE 802.16-2004 standard [1]. This targeted fixed broadband wireless applications via the installation of Customer Premises Equipment (CPE). In December 2005 the IEEE completed the 802.16e-2005 [2] amendment, which added new features to support mobile applications.

Mobile WiMAX extends the original OFDM PHY layer to support multiple-access by using scalable OFDMA [2]. Data streams, to and from individual user equipment, are multiplexed onto groups of subchannels on both the downlink (DL) and uplink (UL). By adopting a scalable PHY architecture, mobile WiMAX is able to support a wide range of bandwidths. The scalability is implemented by allowing the FFT size to vary between 128, 512, 1024, and 2048 to support channel bandwidths of 1.25 MHz, 5 MHz, 10 MHz, and 20 MHz respectively.

Mobile WiMAX supports a full-range of smart antenna techniques; including Alamouti spatial transmit diversity [3], spatial multiplexing [4], and closed-loop MIMO eigen-beamforming (EB) [5].

The WiMAX medium access control (MAC) layer provides a connection-oriented service to the upper layers and ensures that the necessary Quality of Service (QoS) requirements for the MAC Packet Data Unit (PDUs) belonging to different applications are fulfilled. TCP and UDP-based applications have very different QoS criteria. The PHY Packet Error Rate (PER) has a strong impact on the ability to provide QoS to the upper layers. TCP-based applications such as web browsing, email, and FTP may tolerate a relatively high PHY PER due to the use of feedback in the form of end-to-end application ARQs. In contrast, UDP-based applications such as video streaming cannot afford the latency induced by ARQ and therefore require a very low PER at the physical layer.

This paper analyses how the difference in required PHY layer PER results in differences in the achievable performance of a MIMO mobile WiMAX system. It looks at how the SNR threshold between Adaptive Modulation and Coding (AMC) schemes varies with PHY PER. It also looks at how the achievable operating range and peak throughput varies with application type (and hence PHY PER).

The results show a need for cross-layer interaction between the PHY layer and the scheduler in the MAC layer in order to achieve optimum performance. Furthermore it can be seen that because different applications have different QoS requirements, they operate at different performance levels.

The paper is organized as follows: a description of the WiMAX MAC layer and our MIMO DL mobile WiMAX physical layer simulator is given in section II. Different MIMO schemes implemented in our WiMAX simulator are summarised in section III. The underlying wideband channel model is described in section IV. Section V presents the MIMO downlink performance in terms of PER, throughput, and operating range as a function of SNR and link-speed. Finally, performance comparisons between different PHY PER are provided.

II. MOBILE WIMAX DESCRIPTION

A. Medium Access Control (MAC) Layer

The IEEE 802.16e MAC is logically divided into three sublayers: the service-specific convergence sublayer (CS), the common-part sublayer (CPS), and the security sublayer (SS).

The convergence sublayer is responsible for performing all operations that are dependent on the nature of the higher layers. The CS receives higher layer packets known as MAC service data units (SDU). Based on their QoS requirements
each packet (or MAC SDU) is classified into a service flow. In WiMAX a service flow is a MAC transport service provided for transmission of uplink and downlink traffic and is a key concept within the QoS architecture. Each service flow is associated with a unique set of QoS parameters, such as latency, jitter, throughput, and packet error rate, as required by the upper layers. They each have a unique service flow ID (SFID). Service flows are themselves mapped to different scheduling services. The following scheduling services are defined in the WiMAX standard: Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), non-real-time Polling Service (nrtPS) and Best Effort (BE). These are designed for constant bit rate (CBR) real-time, variable bit rate (VBR) real-time, VBR non-real-time, and BE applications, respectively. These scheduling services define the mechanisms used to allocate resources to different applications. Based on its service flow, source address and destination address each packet is assigned a connection ID (CID) which forms a logical connection between the Base Station (BS) and Mobile Station (MS). An MS can have more than one CID at a time, depending on the different applications and their QoS requirements.

The common-part sublayer is independent of the higher-layer protocol and performs operations such as MAC PDU construction, QoS guarantee, bandwidth allocation, and scheduling. MAC SDUs arriving at the CS sublayer are fragmented or concatenated to create MAC PDUs. Once a PDU is constructed, it is handed to the scheduler, which schedules the MAC PDU over the available PHY resource. The scheduler checks the SFID and CID of the MAC PDU in order to determine its QoS requirements. Based on the QoS requirements of the MAC PDUs belonging to different CIDs, the scheduler determines the optimum PHY resource allocation (or bandwidth allocation) for all MAC PDUs, on a frame-by-frame basis. QoS guarantee is managed for each CID individually. Finally, MAC PDUs are encoded to prevent theft of service in the security sublayer [6]. The whole MAC process is illustrated in the Fig. 1.

**B. Physical Layer (PHY)**

The mobile WiMAX standard builds on the principles of OFDM by adopting a Scalable OFDMA-based PHY layer (SOFDMA). SOFDMA supports a wide range of operating bandwidths to flexibly address the need for various spectrum allocation and application requirements. When the operating bandwidth increases, the FFT size is also increased to maintain a fixed subcarrier frequency spacing of 10.94 kHz. This ensures a fixed OFDMA symbol duration. Since the basic resource unit (i.e. the OFDMA symbol duration) is fixed, the impact of bandwidth scaling is minimized to the upper layers. Table 1 shows the relevant parameters for the OFDMA PHY.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT size</td>
<td>128</td>
</tr>
<tr>
<td>Channel bandwidth (MHz)</td>
<td>1.25</td>
</tr>
<tr>
<td>Subcarrier frequency spacing (kHz)</td>
<td>10.94</td>
</tr>
<tr>
<td>Useful symbol period (μs)</td>
<td>91.4</td>
</tr>
<tr>
<td>Guard Time</td>
<td>1/32, 1/16, 1/8, 1/4</td>
</tr>
</tbody>
</table>

TABLE II summarises the OFDMA parameters used in our Mobile WiMAX simulator.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel bandwidth (MHz)</td>
<td>5</td>
</tr>
<tr>
<td>Sampling frequency F_s (MHz)</td>
<td>5.6</td>
</tr>
<tr>
<td>Sampling period T_s (μs)</td>
<td>0.18</td>
</tr>
<tr>
<td>Subcarrier frequency spacing Δ f = F_s/NFFT (kHz)</td>
<td>10.94</td>
</tr>
<tr>
<td>Useful symbol period T_b = 1/Δ f (μs)</td>
<td>91.4</td>
</tr>
<tr>
<td>Guard Time T_g = T_s/K (μs)</td>
<td>11.4</td>
</tr>
<tr>
<td>OFDMA symbol duration T_c = T_b + T_g (μs)</td>
<td>102.9</td>
</tr>
<tr>
<td>Number of used subcarriers (N_used)</td>
<td>421</td>
</tr>
<tr>
<td>Number of pilot subcarriers</td>
<td>60</td>
</tr>
<tr>
<td>Number of data subcarriers</td>
<td>360</td>
</tr>
<tr>
<td>Number of data subcarriers in each subchannel</td>
<td>24</td>
</tr>
<tr>
<td>Number of subchannels</td>
<td>15</td>
</tr>
<tr>
<td>Number of users (N_user)</td>
<td>3</td>
</tr>
<tr>
<td>Number of subchannels allocated to each user (N_user)</td>
<td>5</td>
</tr>
</tbody>
</table>

Fig. 2 shows the block diagram of our MIMO enabled WiMAX simulator. The channel coding stage includes randomization, convolutional encoding (native code rate is 1/2) and puncturing to produce higher code rates. A block interleaver is used to interleave the encoded bits onto separate subcarriers, thus minimizing the impact of burst errors. Once the data has been modulated (using QPSK, 16QAM, or 64QAM) the data is mapped by segmenting the sequence of modulation symbols into a sequence of slots (the minimum data allocation unit) and then mapping these slots into a data region. This is performed such that the lowest numbered slot occupies the lowest numbered logical subchannel among the allocated subchannels. The mapping of the slots continues vertically to the edge of the data region (defined as a group of contiguous logical subchannels), and then moves to the next available OFDMA slot.
After the MIMO processing has been applied, the data symbols are mapped to a data region and assigned to their corresponding logical subcarriers. These logical subcarriers are allocated to physical subcarriers using a specific sub-carrier permutation. Pilots are also inserted at this point together with a guard interval. The final stage is to convert the data into the time-domain.

Our simulation supports a number of link-speeds (see Table III for details). A link-speed is defined by a combination of a modulation scheme and a coding rate. The peak data rate $D$ is calculated as below:

$$D = N_D N_s R_{FEC} R_{STC} / T_s$$  \hspace{1cm} (1)

where $N_D$, $N_s$, $R_{FEC}$, $R_{STC}$ and $T_s$ denote the number of assigned data subcarriers to each user, the bits per sub-carrier, the forward error correction (FEC) coding rate, the space-time coding rate, and the OFDMA symbol duration respectively.

**TABLE III: DOWNLINK MOBILE WIMAX LINK-SPEEDS**

<table>
<thead>
<tr>
<th>Mode (Link-Speed)</th>
<th>No. of coded bits per subchannel DL</th>
<th>No. of data bits per subchannel DL</th>
<th>STBC 2x2 DL bit rate/user (Mbps)</th>
<th>SM 2x2 DL bit rate/user (Mbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QPSK 1/2</td>
<td>48</td>
<td>24</td>
<td>1.17</td>
<td>2.34</td>
</tr>
<tr>
<td>QPSK 3/4</td>
<td>48</td>
<td>36</td>
<td>1.75</td>
<td>3.5</td>
</tr>
<tr>
<td>16 QAM 1/2</td>
<td>96</td>
<td>48</td>
<td>2.33</td>
<td>4.66</td>
</tr>
<tr>
<td>16 QAM 3/4</td>
<td>96</td>
<td>72</td>
<td>3.50</td>
<td>7</td>
</tr>
<tr>
<td>64 QAM 1/2</td>
<td>144</td>
<td>72</td>
<td>3.50</td>
<td>7</td>
</tr>
<tr>
<td>64 QAM 3/4</td>
<td>144</td>
<td>96</td>
<td>4.66</td>
<td>9.33</td>
</tr>
<tr>
<td>64 QAM 5/4</td>
<td>144</td>
<td>108</td>
<td>5.25</td>
<td>10.50</td>
</tr>
</tbody>
</table>

**III. MIMO SCENARIOS DESCRIPTION**

A. **Space-Time Block Coding (STBC)**

Our mobile WiMAX simulator implements the Alamouti scheme [3] on the DL to provide transmit and receive diversity.

B. **Spatial Multiplexing (SM)**

Mobile WiMAX supports SM to increase the peak error-free data rate by transmitting separate data streams from each antenna. For those modes based on SM our simulator uses an MMSE receiver to remove the inter-stream interference on a per sub-carrier basis [4].

C. **Eigen-beamforming**

Eigen-beamforming is a closed-loop MIMO technique in which knowledge of the Channel State Information (CSI) is available at the transmitter. The technique utilises singular-value decomposition (SVD) of the MIMO channel $H$ to transform it into $N$ equivalent parallel SISO eigen-channels [5]. These $N$ SISO channels are used at the transmitter to direct $N$ data sub streams so that they can be extracted without interference at the receiver. The value $N$ is given by the minimum number of antennas at the transmitter ($N_T$) or the receiver ($N_R$).

Two eigen-beamforming schemes are investigated in this paper. The first maximises diversity by restricting the transmission of data symbols to the strongest eigen-mode. In this case all the transmit power is focused on the strongest spatial channel and the weaker spatial modes are unused (thus lowering the peak error-free data rate). This scenario is beneficial in highly correlated channels, where the mean array gain tends to $N_R N_T$ [7]. The second eigen-beamforming approach transmits data on all the spatial channels. This is used to maximise the error-free throughout, and is applicable to highly uncorrelated channels (where the weakest eigen-value is still sufficient for data transmission).

The main disadvantage of eigen-beamforming (compared to the earlier STBC and SM techniques) is the need for accurate CSI at the transmitter [8].

**IV. MIMO WIDEBAND CHANNEL MODEL**

Based on the ETSI 3GPP2 spatial channel model (SCM) [9], an urban micro tapped delay line (TDL) model was generated for use in this analysis. The TDL comprises 6 taps with non-uniform delays. The Mobile Station (MS) velocity is assumed to be 40 km/h. The antenna element separation is half a wavelength. A 3 sector base station (BS) is assumed and the channel was generated for 3 MS. The resulting spatial channel correlation coefficient ($\rho$) is 0.16 (highly uncorrelated). The channel delay spread is 279 ns.

**V. DOWNLINK SIMULATION PERFORMANCE ANALYSIS**

In this section the MIMO mobile WiMAX ($N_T = N_R = 2$) throughput and operating range results are presented using the Mobile WiMAX simulator and channel model described in sections II, III and IV. The BS transmits data simultaneously to 3 MSs, with each sharing a common OFDMA symbol. Perfect channel estimation and synchronisation is assumed. The link throughput for each user is estimated from the PER as follows:

$$R = D \times (1 - \text{PER})$$  \hspace{1cm} (2)

where $R$ represents the peak error-free transmission rate for the chosen link-speed (see section II).

The operating range $d$ (in metres) is derived from (3) using the COST 231 Walfish-Ikegami model [7]

$$P_t(dB) = 55.9 + 38 \log_{10}(d) + (24.5 + 1.5 f_c/825) \log_{10}(f_c)$$  \hspace{1cm} (3)

where $f_c$ is the carrier frequency in MHz. The path loss $P_t$ (dB) is derived from the link budget equation (4).

$$P_t = G_T + G_{AT} + R - P_L$$  \hspace{1cm} (4)

where $P_t$ represents the transmit power (43dBm), $G_T$ and $G_{AT}$ represent the antenna gains of the basestation and mobile terminal respectively (15dBi and 0dBi), and $P_R$ represents the received power which is derived from (5). The values quoted in brackets represent the numbers used in our link-budget.

$$P_R = SNR - N$$  \hspace{1cm} (5)

where $SNR$ is the received signal-to-noise ratio and $N$ is the
thermal noise floor (dBm). This can be calculated as

\[ N = 10\log_{10}(\kappa F) + 10\log_{10}(W) + NF \]  

(6)

where \( \kappa = 1.38 \times 10^{-23} \text{J/Hz/K} \), \( F \) is the temperature in Kelvin (300K), \( W \) is bandwidth in Hz, and \( NF \) is the receiver noise figure (7dB).

In our throughput and operating range calculations we use different PER thresholds for TCP-based and UDP-based applications: 10% PER is considered the highest PER acceptable for TCP-based applications; any PER in excess of this value is assumed to be too severe to maintain a practical data link and is not included in the calculations. For UDP-based applications the PER threshold is set to 1%. Differences in the PER thresholds result in differences in the achievable throughput and operating ranges as well as the SNR thresholds used in the AMC scheme. This is shown in the next section.

A. PER Analysis

Fig. 3 compares the PER performance for each of the MIMO schemes. It can be seen that the PER performance of the SM eigen-beamforming technique (denoted SVD SM in the figure) offers a significant improvement when compared with SM, especially at the high channel coding rate. More specifically, at a PER of \( 10^{-2} \), for 16QAM 1/2 rate and 3/4 rate the performance improvement is 7dB and 2.5dB respectively. Compared to SM, for a coding rate of 3/4, SVD only provides an array gain; meanwhile for the 1/2 rate code the SVD scheme provides a much higher diversity order (equivalent to STBC). This agrees with [5] where the maximum diversity order of a coded SVD scheme is \( N_R \times N_T \).

B. Throughput Analysis

Fig. 4 shows the throughput versus SNR envelopes of the SM MIMO mode for PER thresholds of 10% and 1% respectively. The envelope was generated using adaptive modulation and coding (AMC) to increase and/or decrease the link-speed depending on the received SNR while still maintaining the appropriate QoS threshold PER.

It can be seen from fig. 4 that in order to offer a useable connection the 1% PER case needs a minimum SNR of 14dB. This is 4dB higher than the 10% PER case (10dB). The reason for this is that at SNR values lower than 14dB, the corresponding PER is higher than 1% and therefore no throughput can be achieved. It is also clear that in the 1% PER case a higher SNR is needed in order to achieve the higher link-speed supported by the 10% PER case. When using a 1% PER limit the system often operates at a lower link-speed (compared to the 10% PER threshold). For example, in the SNR range from 17dB to 20dB the 10% PER case can select the 16QAM 1/2 rate link-speed, meanwhile the 1% PER case can only operate at the lower QPSK 1/2 rate link-speed. This significantly decreases the throughput of the system (but reduces the PHY PER). When an MS has two applications (one TCP and one UDP) running at the same time, the UDP application will often receive a lower throughput than the TCP application.

Fig. 5 illustrates the throughput versus SNR envelopes of the STBC MIMO mode for PER thresholds of 10% and 1%. For SNR values below 18dB we observe the same trends as before. Firstly a higher minimum SNR is required for the lowest link-speed and secondly lower link-speeds are available within the same SNR range for the 1% PER threshold (compared to the 10% PER threshold). However, at SNR values above 18 dB, both cases achieve the same maximum throughput when operating at the same link-speed and in the same SNR range. The reason for this is that STBC provides a significant diversity and array gain when compared to SM. This allows both applications to operate at low PER, even at the highest link-speed (64QAM 3/4 rate) when the SNR exceeds a value of 18dB.
Fig. 6 and Fig. 7 show the throughput versus SNR envelopes of the SVD-SM and SVD-DE MIMO modes for PER thresholds of 10% and 1% respectively. In order to switch to the same higher link-speed, SVD-SM and SVD-DE require lower SNR values than their counterpart open-loop MIMO SM and STBC schemes. This is due to the improvement gained by utilising CSI at the transmitter.

C. Operating Range Analysis

Fig. 8 illustrates the throughput versus distance envelopes of the SM MIMO mode for PER thresholds of 10% and 1%. The envelope is also generated using AMC as in section B. We can see that the 10% PER case can achieve a higher throughput and a greater operating range. This is because at certain ranges the 1% PER case has to operate at a lower link-speed, resulting in a lower throughput. It can also be seen that the maximum achievable operating range of the 10% PER case is higher than the 1% PER case (727m rather than 557m). Finally we can see that in the 10% PER case the maximum throughput is achieved at a range of 250m. In contrast, in the 1% PER case the maximum throughput is only 7 Mbps even when the MS is very close to the BS. This is much less than the maximum system capacity of 10Mbps.

Fig. 9 shows the throughput versus distance envelopes for the STBC MIMO mode at PER thresholds of 10% and 1%. The only difference between this and the SM mode is that at short ranges (up to 450m) both the 10% and 1% case can operate at the same link-speed to achieve maximum throughput. The same trends are illustrated in Fig. 10 and Fig. 11 for the SVD-SM and SVD-DE MIMO modes. Again we observe that the eigen-MIMO modes achieve a better operating range than either SM (with MMSE) or STBC.
VI. CONCLUSIONS

This paper has presented a detailed study of the throughput and operating range of a MIMO mobile WiMAX system under two different PHY PER QoS thresholds. The MIMO channel was modelled using the 3GPP spatial channel model. The simulation is fully compliant with the 802.16e-2006 standard. Throughput and operating range results were presented for a number of MIMO modes and all standardised link-speeds. The results clearly show that TCP applications achieve a higher throughput and operating range when compared with UDP applications. This occurs because of the higher PER threshold tolerated by the former protocol. Applications such as FTP and web-browsing can operate over larger distances than those sensitive to PER, such as VoIP or video streaming. This is important in the cell planning process, where it is critical to support the QoS requirements of multiple applications running on multiple MSs.

Another important observation is that the AMC SNR switching points are dependent on the application QoS requirements. UDP applications require higher SNR in order to switch to the same link-speed as TCP applications. This demonstrates the importance of cross-layer interaction when determining the AMC switching points. Overall, the system needs to know about the application’s QoS requirements as well as the received SNR in order to choose the appropriate switching point and link-speed for the flow.

REFERENCES