An Efficient Analytical Model for the Dimensioning of WiMAX Networks

Bruno Baynat[1], Georges Nogueira[1], Masood Maqbool[2], and Marceau Coupechoux[2]

[1] Universite Pierre et Marie Curie - Paris, France {`firstname.lastname`}`@lip6.fr`
[2] Telecom ParisTech - Paris, France {`firstname.lastname`}`@telecom-paristech.fr`

Abstract. This paper tackles the challenging task of developing a simple and accurate analytical model for performance evaluation of WiMAX networks. The need for accurate and fast-computing tools is of primary importance to face complex and exhaustive dimensioning issues for this promising access technology. In this paper, we present a generic Markovian model developed for three usual scheduling policies (slot sharing fairness, throughput fairness and opportunistic scheduling) that provides closed-form expressions for all the required performance parameters at a click speed. This model is compared in depth with realistic simulations that show its accuracy and robustness regarding the different modeling assumptions. Finally, the speed of our analytical tool allows us to carry on dimensioning studies that require several thousands of evaluations, which would not be tractable with any simulation tool.

Keywords: WiMAX, performance evaluation, dimensioning, analytical models

1 Introduction

The evolution of last-mile infrastructure for wired broadband networks faces acute implications such as difficult terrain and high cost-to-serve ratio. Latest developments in wireless domain could not only address these issues but could also complement the existing framework. One of such highly anticipated technologies is WiMAX (Worldwide Interoperability for Microwave Access) based on IEEE standard 802.16. The first operative version of IEEE 802.16 is 802.16-2004 (fixed/nomadic WiMAX) [1]. It was followed by a ratification of mobile WiMAX amendment IEEE 802.16e in 2005 [2]. On the other hand, the consortium WiMAX Forum was found to specify profiles (technology options are chosen among those proposed by the IEEE standard), define an end-to-end architecture (IEEE does not go beyond physical and MAC layer), and certify products (through inter-operability tests).

Some WiMAX networks are already deployed but most operators are still under trial phases. As deployment is coming, the need arises for manufacturers and operators to have fast and efficient tools for network design and performance evaluation. In [3] authors propose an analytical model for studying the

random access scheme of IEEE 802.16d. Niyato and Hossain [4] formulate the bandwidth allocation of multiple services with different QoS requirements by using linear programming. They also propose performance analysis, first at connection level, and then, at packet level. In the former case, variations of the radio channel are however not taken into account. In the latter case, the computation of performance measures rely on multi-dimensional Markovian model that requires numerical resolutions. Not specific to WiMAX systems, generic analytical models for performance evaluation of cellular networks with varying channel conditions have been proposed in [5–7]. The models presented in these articles are mostly based on multi-class processor-sharing queues with each class corresponding to users having similar radio conditions and subsequently equal data rates. The variability of radio channel conditions at flow level is taken into account by integrating propagation models, mobility models or spatial distribution of users in a cell. In order to use classical PS-queues results, these papers consider implicitly that users can only switch class between two successive data transfers. However, as highlighted in the next section, in WiMAX systems, radio conditions and thus data rates of a particular user can change frequently during a data transfer. In addition, capacity of a WiMAX cell may vary as a result of varying radio conditions of users. As a consequence, any PS, DPS (discriminatory PS) or even GPS (generalized PS) queue is not appropriate for modeling these channel variations.

In this paper, we develop a novel and generic analytical model that takes into account frame structure, precise slot sharing-based scheduling and channel quality variation of WiMAX systems. Unlike existing models [5–7], our model is adapted to WiMAX systems' assumptions and is generic enough to integrate any appropriate scheduling policy. Here, we consider three classical policies: *slot sharing fairness*, *instantaneous throughput fairness*, and *opportunistic*. For each of them, we develop closed-form expressions for all performance metrics. Moreover, our approach makes it possible to take into account the so-called "outage" situation. A user experiences an outage, if at a given time radio conditions are so bad that it cannot transfer any data and is thus not scheduled. Once again, classical PS-like queues are not appropriate to model this feature.

The paper is organized as follows. Modeling assumptions are presented in Section 2. Section 3 presents the generic analytical model and its adaption to the three considered scheduling policies. Validation and robustness are discussed in Section 4. Section 5 finally gives an example of WiMAX dimensioning process.

2 Modeling assumptions

The development of our analytical model is based on several assumptions related to the system, the channel, the traffic and the scheduling algorithm. We present here these assumptions. All of them will be discussed in Section 3.4, and, as will be developed in that section, most of them can be relaxed, if necessary, by slightly modifying the model. Wherever required, related details of WiMAX system are specified. Various notations are also introduced in this section.

A WiMAX time division duplex (TDD) frame comprises of slots that are the smallest unit of resource and which occupies space both in time and frequency domain. A part of the frame is used for overhead (e.g., DL_MAP and UL_MAP) and the rest for user data. The duration T_F of this TDD frame is equal to 5 ms [2].

System assumptions We consider a single WiMAX cell and focus on the downlink part which is a critical portion of asymmetric data traffic.

1. Overhead in the TDD frame is assumed to be constant and independent of the number of concurrent active mobile station (MS). As a consequence, the total number of slots available for data transmission in the downlink part is constant and will be denoted by N_S.
2. We assume that the number of MS that can simultaneously be in active transfer is not limited. As a consequence, any connection demand will be accepted and no blocking can occur.

One of the important features of IEEE 802.16e is link adaptation: different modulation and coding schemes (MCS) allows a dynamic adaptation of the transmission to the radio conditions. As the number of data subcarriers per slot is the same for all permutation schemes, the number of bits carried by a slot for a given MCS is constant. The selection of appropriate MCS is carried out according to the value of signal to interference plus noise ratio (SINR). In case of outage, i.e., if the SINR is too low, no data can be transmitted without error. We denote the radio channel states as: MCS_k, $1 \leq k \leq K$, where K is the number of MCS. By extension, MCS_0 represents the outage state. The number of bits transmitted per slot by a MS using MCS_k is denoted by m_k. For the particular case of outage, $m_0 = 0$.

Channel assumption The MCS used by a given MS can change very often because of the high variability of the radio link quality.

3. We assume that each MS sends a feedback channel estimation on a frame by frame basis, and thus, the base station (BS) can change its MCS every frame. Since we do not make any distinction between users and consider all MS as statistically identical, we associate a probability p_k with each coding scheme MCS_k, and assume that, at each time-step T_F, any MS has a probability p_k to use MCS_k.

Traffic assumptions The traffic model is based on the following assumptions.

4. All users have the same traffic characteristics. In addition, we don't consider any QoS differentiation here.
5. We assume that there is a fixed number N of MS that are sharing the available bandwidth of the cell.

6. Each of the N MS is assumed to generate an infinite length ON/OFF elastic traffic. An ON period corresponds to the download of an element (e.g., a web page including all embedded objects). The downloading duration depends on the system load and the radio link quality, so ON periods must be characterized by their size. An OFF period corresponds to the reading time of the last downloaded element, and is independent of the system load. As opposed to ON, OFF periods must then be characterized by their duration.
7. We assume that both ON sizes and OFF durations are exponentially distributed. We denote by $\bar{x}_{on}$ the average size of ON data volumes (in bits) and by $\bar{t}_{off}$ the average duration of OFF periods (in seconds).

Scheduling assumption The scheduling algorithm is responsible for allocating radio resources to users. In wireless networks, scheduling may take into account their radio link quality. In this paper, we have considered three traditional schemes. The slot fairness scheduling allocates the same number of slots to all active users. The throughput fairness scheduling ensures that all active users have the same instantaneous throughput. The opportunistic scheduling gives all resources to active users with the best channel.

8. At any time and for all scheduling policies, if there is only one active user, we assume that the scheduler can allocate all the available slots for its transfer.

3 WiMAX Analytical model

3.1 Markovian model

A first attempt for modeling this system would be to develop a multi-dimensional Continuous Time Markov Chain (CTMC). A state $(n_0, ..., n_K)$ of this chain would be a precise description of the current number n_k of MS using coding scheme MCS_k, $0 \leq k \leq K$ (including outage). The derivation of the transitions of such a model is an easy task. However the complexity of the resolution of this model makes it intractable for any realistic value of K. In order to work around the complexity problem, we aggregate the state description of the system into a single dimension n, representing the total number of concurrent active MS, regardless of the MCS they use. The resulting CTMC is thus made of $N + 1$ states as shown in Fig 1.

– A transition out of a generic state n to a state $n + 1$ occurs when a MS in OFF period starts its transfer. This "arrival" transition corresponds to one MS among the $(N - n)$ in OFF period, ending its reading, and is performed with a rate $(N - n)\lambda$, where λ is defined as the inverse of the average reading time: $\lambda = \frac{1}{\bar{t}_{off}}$.
– A transition out of a generic state n to a state $n - 1$ occurs when a MS in ON period completes its transfer. This "departure" transition is performed with a generic rate $\mu(n)$ corresponding to the total departure rate of the frame when n MS are active.
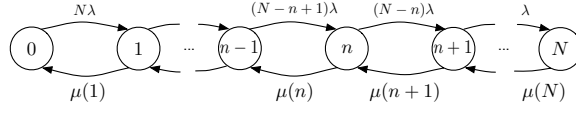
Fig. 1. General CTMC with state-dependent departure rates.

Obviously, the main difficulty of the model resides in estimating the aggregate departure rates $\mu(n)$. In order to do so, we first express $\mu(n)$ as follows:

$$\mu(n) = \frac{\bar{m}(n)\, N_S}{\bar{x}_{on}\, T_F}, \tag{1}$$

where $\bar{m}(n)$ is the average number of bits per slot when there are n concurrent active transfers. Obviously, $\bar{m}(n)$ depends on K, the number of MCS, and p_k, $0 \le k \le K$, the MCS vector probability. It also strongly depends on n, because the number of bits per slot must be estimated by considering all possible distributions of the n MS between the $K+1$ possible MCS (including outage). It is worthwhile noting that the parameters $\bar{m}(n)$ finally depend on the scheduling policy, as it defines, at each time-step, the quantity of slots given to each of the n MS with respect to the MCS they use.

In order to provide a generic expression of $\bar{m}(n)$, we define $x_k(j_0, ..., j_K)$ the proportion of the resource (i.e., of the N_S slots) that is associated to a MS using MCS_k, when the current distribution of the n MS among the $K+1$ coding schemes is $(j_0, ..., j_K)$. The average number of bits per slot, $\bar{m}(n)$, when there are n active users, can then be expressed as follows:

$$\bar{m}(n) = \sum_{\substack{(j_0, ..., j_K) = (0, ..., 0)| \\ j_0 + ... + j_K = n \\ j_0 \neq n}}^{(n, ..., n)} \left(\sum_{k=1}^{K} m_k j_k x_k(j_0, ..., j_K) \right) \binom{n}{j_1, ..., j_K} \prod_{k=0}^{K} p_k^{j_k}, \tag{2}$$

where $\prod_{k=0}^{K} p_k^{j_k}$ is the probability of any distribution of the n MS such that the number of MS using MSC_k is j_k, and $\binom{n}{j_0, ..., j_K}$ is the multinomial coefficient that takes into account all such possibles distributions.

3.2 Scheduling policy modeling

We now present the adaptation of the model, for the three specific scheduling policies we consider in this paper. For each of them we provide closed-form expressions for the average number of bits per slots, $\bar{m}(n)$.

Slot sharing fairness Each time-step, the scheduler equally shares the N_S slots among the active users that are not in outage. If, at a given time-step, there are n active MS, each of the MS that are not in outage receives a portion $\frac{N_S}{n-j_0}$ of the whole resource. As a consequence, the proportion of the resource that

is associated to a MS using MCS_k, is thus given by: $x_k(j_0, ..., j_K) = \frac{1}{n-j_0}$ for any $k \neq 0$. By replacing these proportions in generic expression (2) we obtain:

$$\bar{m}(n) = \sum_{\substack{(j_0, ..., j_K) = (0, ..., 0)| \\ j_0 + ... + j_K = n \\ j_0 \neq n}}^{(n, ..., n)} \frac{n!}{n - j_0} \left(\sum_{k=1}^{K} m_k j_k \right) \prod_{k=0}^{K} \frac{p_k^{j_k}}{j_k!}. \tag{3}$$

Instantaneous throughput fairness The resource is shared in order to provide the same instantaneous throughput to all active users that are not in outage. This policy allows MS using MCS with a low bit rate per slot to obtain, at a given time-step, proportionally more slots compared to MS using a MCS with a high bit rate per slot. In order to respect instantaneous throughput fairness between all active users that are not in outage, the $x_k(j_0, ..., j_K)$ must be such that: $m_k x_k(j_0, ..., j_K) = C$ for any $k \neq 0$, where C is a constant such that $\sum_{k=1}^{K} j_k x_k(j_0, ..., j_K) = 1$. By replacing the proportions $x_k(j_0, ..., j_K)$ in generic expression (2), the average number of bits per slot $\bar{m}(n)$ becomes:

$$\bar{m}(n) = \sum_{\substack{(j_0, ..., j_K) = (0, ..., 0)| \\ j_0 + ... + j_K = n \\ j_0 \neq n}}^{(n, ..., n)} \frac{(n - j_0) \, n! \prod_{k=0}^{K} \frac{p_k^{j_k}}{j_k!}}{\sum_{k=1}^{K} \frac{j_k}{m_k}}. \tag{4}$$

Opportunistic scheduling All the resource is given to users having the highest transmission bit rate, i.e., the better radio conditions and then the better MCS. Without loss of generality, we assume here that the MCS are classified in increasing order: $m_0 < m_1 < ... < m_K$. And even if it is still possible to derive the average bit rates from generic expression (2), we prefer to give here a more intuitive and equivalent derivation.

We consider a system with n current active MS. We denote by $\alpha_i(n)$ the probability of having at least one active user (among n) using MCS_i and none using a MCS giving higher transmission rates (i.e., MCS_j with $j > i$). As a matter of fact, $\alpha_i(n)$ corresponds to the probability that the scheduler gives at a given time-step all the resource to MS that use MCS_i. As a consequence, we can express the average number of bits per slot when there are n active users as:

$$\bar{m}(n) = \sum_{i=1}^{K} \alpha_i(n) m_i. \tag{5}$$

In order to calculate the $\alpha_i(n)$, we first express the probability that there are no MS using a MCS higher than MCS_i as: $p_{\leq i}(n) = \left(1 - \sum_{j=i+1}^{K} p_j \right)^n$. Then, we calculate the probability that there is at least one MS using MCS_i conditioned by the fact that there are no MS using a better MCS: $p_{=i}(n) = 1 - \left(1 - \frac{p_i}{\sum_{j=0}^{i} p_j} \right)^n$. $\alpha_i(n)$ can thus be expressed as: $\alpha_i(n) = p_{=i}(n) \, p_{\leq i}(n)$.

3.3 Performance parameters

The steady-state probabilities $\pi(n)$ can easily be derived from the birth-and-death structure of the Markov chain (depicted in Fig. 1):

$$\pi(n) = \frac{N!}{(N-n)!} \frac{\dfrac{T_F^n \, \rho^n}{n}}{N_S^n \displaystyle\prod_{i=1}^{n} \bar{m}(i)} \pi(0), \tag{6}$$

where ρ is given by relation (7) and plays a role equivalent to the "traffic intensity" of Erlang laws [8], and $\pi(0)$ is obtained by normalization.

$$\rho = \frac{\bar{x}_{on}}{\bar{t}_{off}} \tag{7}$$

The performance parameters of this system can be derived from the steady-state probabilities as follows. The average utilization $\bar{U}$ of the TDD frame is:

$$\bar{U} = \sum_{n=1}^{N} (1 - p_0^n)\pi(n). \tag{8}$$

The average number of active users $\bar{Q}$ is expressed as:

$$\bar{Q} = \sum_{n=1}^{N} n \, \pi(n). \tag{9}$$

The mean number of departures $\bar{D}$ (MS completing their transfer) by unit of time, is obtained as: $\bar{D} = \sum_{n=1}^{N} \pi(n)\,\mu(n)$. From Little's law, we can derive the average duration $\bar{t}_{on}$ of an ON period (duration of an active transfer): $\bar{t}_{on} = \frac{\bar{Q}}{\bar{D}}$. We finally compute the average throughput $\bar{X}$ obtained by each MS in active transfer as:

$$\bar{X} = \frac{\bar{x}_{on}}{\bar{t}_{on}}. \tag{10}$$

3.4 Discussion of the modeling assumptions

Our Markovian model is based on several assumptions presented in Section 2. We now discuss these assumptions one by one (item numbers are related to the corresponding assumptions), evaluate their accuracy, and provide, if necessary and possible, extensions and generalization propositions.

1. DL_MAP and UL_MAP are located in the downlink part of the TDD frame. They contain the information elements that allow MS to identify the slots to be used. The size of these MAPs, and as a consequence the number N_S of available slots for downlink data transmissions, depends on the number of MS scheduled in the TDD frame. In order to relax assumption 1, we can express the number of data slots, $N_S(n)$, as a function of n, the number of active users. This dependency can be easily integrated in the model by replacing N_S by $\prod_{i=1}^{n} N_S(n)$ in relation (6), and N_S by $N_S(n)$ in relation (1).

2. A limit n_{max} on the total number of MS that can simultaneously be in active transfer, can be introduced easily if required. The corresponding Markov chain (Fig. 1) has just to be truncated to this limiting state (i.e., the last state becomes $\min(n_{max}, N)$). As a result, a blocking can occur when a new transfer demand arrives and the limit is reached. The blocking probability can be derived easily from the Markov chain [9].

3. Radio channel may be highly variable or may vary with some memory. Our analytical model only depends upon stationary probabilities of different MCS whatever be the radio channel dynamics. This approach is authenticated through simulations in Section 4.

4. More complex systems with multiple-traffic or differentiation between users would naturally result into more complex models. This is left for future work.

5. Poisson processes are currently used in the case of a large population of users, assuming independence between the arrivals and the current population of the system. As we focus in this paper on the performance of a single cell system, the potential population of users is relatively small. The higher the number of on-going data connections, the less likely the arrival of new ones. Poisson processes are thus a non-relevant choice for our models. Note however that if Poisson assumptions have to be made for connection demand arrivals, one can directly modify the arrival rates of the Markov chain (i.e., replace the state-dependent rates $(N - n)\lambda$ by some constant value, and limit the number of states of the Markov chain as explained above in point 2).

6. Each MS is supposed to generate infinite length ON/OFF session traffic. In [10], an extension to finite length sessions is proposed in the context of (E)GPRS networks, where each MS generates ON/OFF traffic during a session and does not generate any traffic during an inter-session. This work shows that a very simple transformation of traffic characteristics that increases OFF periods by a portion of the inter-session period, enables to derive the average performance from the infinite length session model. The accuracy of this transformation is related to the insensibility of the average performance parameters with regards to the traffic distributions (see next point). A similar transformation can be applied to our WiMAX traffic model.

7. Memoryless traffic distributions are strong assumptions that are validated by several theoretical results on PS-like queues. Several works on insensitivity have shown that the average performance parameters are insensitive to the distribution of ON and OFF periods [11–13]. In its generic form, our model is no longer equivalent to any PS-like queue, but we show in Section 4 by comparing our model to extensive simulations (using Pareto distributions), that insensibility still holds or is at least a very good approximation.

8. In some cellular networks (e.g. (E)GPRS), MS have limited transmission capabilities because of hardware considerations. This constraint defines a maximum throughput the network interface can reach or a maximum number of resource units that can be used by the MS. This characteristic has been introduced in the case of (E)GPRS networks [9] and consists in reducing the departure rates of the first states of the Markov chain. The same idea can be applied to our WiMAX model.

4 Validation

In this section we discuss the validation and robustness of our analytical model through extensive simulations. For this purpose, a simulator has been developed that implements an ON/OFF traffic generator and a wireless channel for each user, and a centralized scheduler that allocates radio resources, i.e., slots, to active users on a frame by frame basis.

4.1 Simulation Models

System Parameters System bandwidth is assumed to be 10 MHz. The downlink/uplink ratio of the WiMAX TDD frame is considered to be 2/3. We assume for the sake of simplicity that the protocol overhead is of fixed length (2 symbols). Considering subcarrier permutation PUSC, the total number of data slots (excluding overhead) per TDD downlink sub-frame is $N_S = 450$.

Traffic Parameters In our analytical model, we consider an elastic ON/OFF traffic. Mean values of ON data volume (main page and embedded objects) and OFF period (reading time), are 3 Mbits and 3 s respectively.

In the first phase (validation study), we assume that the ON data volume is exponentially distributed as it is the case in the analytical model assumptions. Although well adapted to Markov theory based analysis, exponential law does not always fit the reality for data traffic. This is the reason why we consider truncated Pareto distributions in the second phase (the robustness study). Recall that the mean value of the truncated Pareto distribution is given by equation $\bar{x}_{on} = \frac{\alpha b}{\alpha - 1} \left[1 - (b/q)^{\alpha - 1} \right]$, where α is the shape parameter, b is the minimum value of Pareto variable and q is the cutoff value for truncated Pareto distribution. Two values of q are considered: lower and higher. The mean value in both cases ($q = 300$ Mbits and $b = 611822$ bits for the higher cutoff and $q = 3000$ Mbits and $b = 712926$ bits lower cutoff) is 3 Mbits for the sake of comparison with the exponential model. The value of $\alpha = 1.2$ has been adopted from [14].

Channel Models A generic method for describing the channel between the BS and a MS is to model the transitions between MCS by a finite state Markov chain (FSMC). The chain is discrete time and transitions occurs every L frames, with $L\,T_F < \bar{t}_{coh}$, the coherence time of the channel. In our case, and for the sake of simplicity, $L = 1$. Such a FSMC is fully characterized by its transition matrix $P_T = (p_{ij})_{0 \leq i,j \leq K}$, where state 0 represents outage. Stationary probabilities p_k provide the long term probabilities for a MS to receive data with MCS k.

In our analytical study, channel model is assumed to be memoryless, i.e., MCS are independently drawn from frame to frame for each user, and the discrete distribution is given by the $(p_i)_{0 \leq i,j \leq K}$. This corresponds to the case where $p_{ij} = p_j$ for all i. This simple approach, referred as the *memoryless channel model*, is the one considered in the validation study. Let $P_T(0)$ be the transition matrix associated to the memoryless model.

Table 1. Stationary probabilities.

Channel model	Memoryless	Average	Combined	
			good 50% MS	bad 50% MS
a	0	0.5	0.5	0.5
p_0	0.225	0.225	0.020	0.430
p_1	0.110	0.110	0.040	0.180
p_2	0.070	0.070	0.050	0.090
p_3	0.125	0.125	0.140	0.110
p_4	0.470	0.470	0.750	0.190

Table 2. Channel parameters.

Channel state $\{0, ..., K\}$	MCS and outage	Bits per slot m_k
0	Outage	$m_0 = 0$
1	QPSK-1/2	$m_1 = 48$
2	QPSK-3/4	$m_2 = 72$
3	16QAM-1/2	$m_3 = 96$
4	16QAM-3/4	$m_4 = 144$

In the robustness study, we introduce two additional channel models with memory. In these models, the MCS observed for a given MS in a frame depends on the MCS observed in the previous frame according to the FSMC presented above. The transition matrix is derived from equation $P_T(a) = aI + (1-a)P_T(0)$ given that $0 \leq a \leq 1$. In this equation, I is the identity matrix and parameter a is a measure of the channel memory. A MS maintains its MCS for a certain duration with mean $\bar{t}_{coh} = 1/(1-a)$. With $a = 0$, the transition process becomes memoryless. On the other extreme, with $a = 1$, the transition process will have infinite memory and MS will never change its MCS. For simulations we have taken a equal to 0.5, so that the channel is constant in average 2 frames. This value is consistent with the coherence time given in [15] for 45 Km/h at 2.5 GHz. We call the case where all MS have the same channel model with memory ($a = 0.5$), the *average channel model*. Note that the stationary probabilities of the average channel model are the same as those of the memoryless model.

As the channel depends on the BS-MS link, it is possible to refine the previous approach by considering part of the MS to be in a "bad" state, and the rest in a "good" state. Bad and good states are characterized by different stationary probabilities but have the same coherence time. In the so called *combined channel model*, half of the MS are in a good state, the rest in a bad state, and a is kept to 0.5 for both populations.

Three models are thus considered: the memoryless, the average, and the combined channel models. Wireless channel parameters are summarized in Tab. 2. Considered MCS are given including outage, and for each of them, the number of bits transmitted per slot. Channel stationary probabilities are given in Tab. 1. The probabilities for the combined model are obtained by averaging corresponding values of good and bad model stationary probabilities.

4.2 Simulation Results

In this section, we first present a comparison between the results obtained through our analytical model and scheduling simulator. The output parameters in consideration are $\bar{U}$, $\bar{X}$, and $\pi(n)$ (see Section 3.3).

(a) Average utilization.　(b) Average throughput per (c) Stationary probabilities
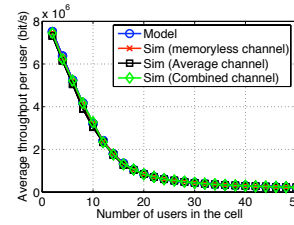　　　　　　　　　　　　user.　　　　　　　　　　for $N = 50$.
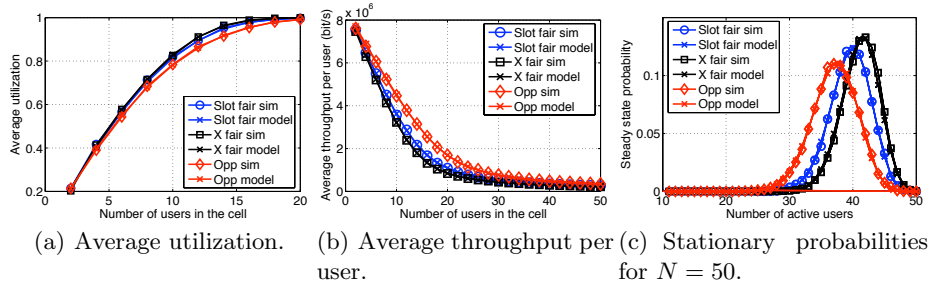
Fig. 2. Validation for the three scheduling policies with $\bar{x}_{on} = 3$ Mbits and $\bar{t}_{off} = 3$ s.
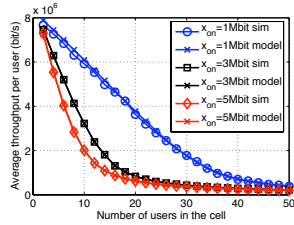
Fig. 3. Average throughput per user for different loads.
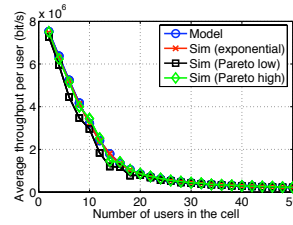
Fig. 4. Average throughput per user for different traffic distributions.

Fig. 5. Average throughput per user for different channel models.

Validation Study In this study, simulations take into account the same traffic and channel assumptions as those of the analytical model. However, in simulator MCS of users are determined on per frame basis and scheduling is carried out in real time, based on MCS at that instant. The analytical model on the other hand, considers stationary probabilities of MCS only.

Fig. 2(a, b) show respectively the average channel utilization ($\bar{U}$) and the average instantaneous throughput per user ($\bar{X}$) for the three scheduling schemes. It is clear that simulation and analytical results show a good agreement: for both utilization and throughput, the maximum relative error stays below 6% and the average relative error is less than 1%. Fig. 2(c) further proves that our analytical model is a very good description of the system: stationary probabilities $\pi(n)$ are compared with those of simulations for a given total number $N = 50$ of MS. Again results show a perfect match between two methods with an average relative error below 9%. At the end, Fig. 3 shows the validation for three different loads (1, 3 and 5 Mbps). Our model shows a comparable accuracy for all three load conditions with a maximum relative error of about 5%.

Robustness Study In order to check the robustness of our analytical model towards distribution of ON data volumes, simulations are carried out for exponential and truncated pareto (with lower and higher cutoff). The results for

this analysis are shown in Fig. 4. The average relative error between analytical results and simulations stays below 10% for all sets. It is clear that considering a truncated Pareto distribution has little influence on the design parameters.

Next we evaluate the robustness of our analytical model with respect to the channel model. We compare the analytical results with simulation for the three pre-cited channel models: memoryless, average and combined (with stationary probabilities given in Tab. 1). If we look at the plot of Fig. 5, we can say that even for a complex wireless channel, our analytical model shows considerable robustness with an average relative error below 7%. We can thus deduce that for designing a WiMAX network, channel information is almost completely included in the stationary probabilities of the MCS.

5 Network design

In this section we provide some examples to demonstrate application of our model while considering throughput fairness scheduling. However, results can be obtained in the same manner for other scheduling schemes by using their respective average bits per slot $\bar{m}(n)$.

5.1 Performance graphs

We first draw 3-dimensional surfaces where performance parameters are function of, e.g., N, the number of users in the cell and ρ, the combination of traffic parameters. For each performance parameter, the surface is cut out into level lines and the resulting 2-dimensional projections are drawn. The step between level lines can be arbitrarily chosen.

The average radio resource utilization of the WiMAX cell $\bar{U}$, and the average throughput per user $\bar{X}$ for any MS in the system are presented in Fig. 6 and 7 (corresponding to the radio link characteristics presented in Section 4). These graphs allow to directly derive any performance parameter knowing the traffic load profile, i.e., the couple (N, ρ). Each graph is the result of several thousands of input parameter sets. Obviously, any simulation tool or even any multi-dimensional Markov chain requiring numerical resolution, would have precluded the drawing of such graphs.

5.2 Dimensioning study

In this section, we show how our model can be advantageously used for dimensioning issues. Two examples, each respecting a certain QoS criterion, are given.

In Fig. 8 we find minimum number N of MS in the cell to guarantee that the average radio utilization is over 50%. This kind of criterion allows operators to maximize the utilization of network resource in comparison with the traffic load of their customers. For a given traffic load profile and a given set of system parameters, the point of coordinates (N_S, ρ) in the graph is located between two level lines, and the level line with the higher value gives the optimal value of N.
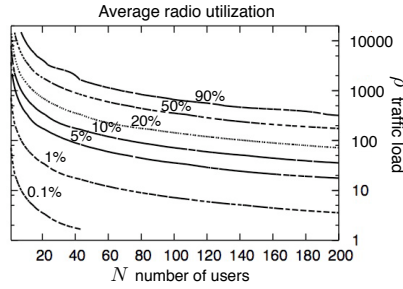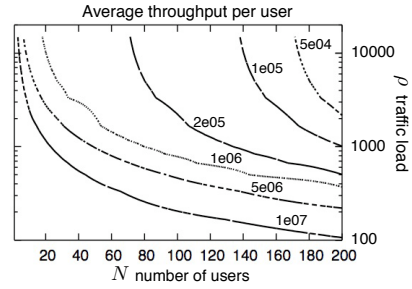
Fig. 6. Average utilization $\bar{U}$.

Fig. 7. Average throughput per user $\bar{X}$.
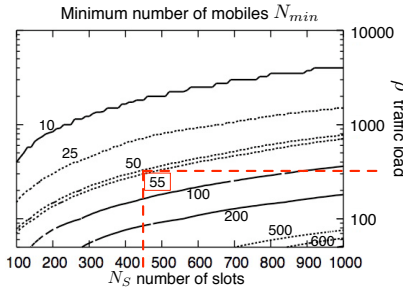

Fig. 8. Dimensioning the minimum value of N for having $\bar{U} \geq 50\%$.
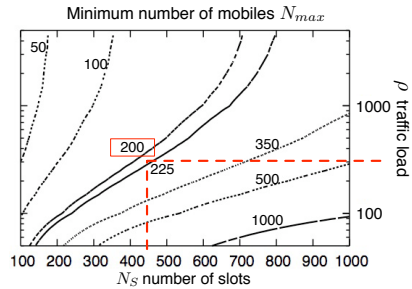

Fig. 9. Dimensioning the maximum value of N for having $\bar{X} \geq 50$ Kbps per user.

The QoS criterion chosen for second example is the user throughput. We have taken 50 Kbps, an arbitrary value of minimum user throughput. Next we find the maximum number N_{max} of users in the cell to guarantee the minimum throughput threshold. In Fig. 9, a given point (N_S, ρ) is located between two level lines. The line with the lower value gives N_{max}. As explained before, the average throughput per user is inversely proportional to N.

The graphs of Fig. 9 and 8 can be jointly used to satisfy multiple QoS criteria. For example, if we have a WiMAX cell configured to have $N_S = 450$ slots and a traffic profile given by $\rho = 300$ (e.g., $x_{on} = 1.2$ Mbits and $t_{off} = 20$ s), Fig. 8 gives $N_{min} = 55$, and Fig. 9 gives $N_{max} = 200$. The combination of these two graphs recommend to have a number of users $N \in [55; 200]$ to guarantee a reasonable resource utilization and a minimum throughput to users.

6 Conclusion

As deployment of WiMAX networks is underway, need arises for operators and manufacturers to develop dimensioning tools. In this paper, we have presented novel analytical models for WiMAX networks and elastic ON/OFF traffic. The models are able to derive Erlang-like performance parameters such as throughput per user or channel utilization. Based on a one-dimensional Markov chain and

the derivation of average bit rates, whose expressions are given for three main scheduling policies (slot fairness, throughput fairness and opportunistic scheduling), our model is remarkably simple. The resolution of model provides closed-form expressions for all the required performance parameters at a click-speed. Extensive simulations have validated the model's assumptions. The accuracy of the model is illustrated by the fact that, for all simulation results, maximum relative errors do not exceed 10%. Even if the traffic and channel assumptions are relaxed, analytical results still match very well with simulations that shows the robust nature of our model.

References

1. Group, I..S.W.: IEEE Standard for local and metropolitan area networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems (2004)
2. Group, I..S.W.: Draft IEEE std 802.16e/D9. IEEE Standard for local and metropolitan area networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems. (2005)
3. Vinel, A., Zhang, Y., Lott, M., Tiurlikov, A.: Performance analysis of the random access in IEEE 802.16. In: Proc. of IEEE PIMRC, p.1596–1600. (September 2005)
4. Niyato, D., Hossain, E.: A queuing-theoretic and optimization-based model for radio resource management in IEEE 802.16 broadband networks. IEEE ToC (vol. 55) (2006)
5. Borst, S.: User-level performance of channel-aware scheduling algorithms in wireless data networks. In: IEEE Infocom. (2003)
6. Bonald, T., Proutiere, A.: Wireless downlink channels: User performance and cell dimensioning. In: ACM Mobicom. (2003)
7. Liu, S., Virtamo, J.: Performance Analysis of Wireless Data Systems with a Finite Population of Mobile Users. In: 19th ITC. (2005)
8. Engset, T.O.: On the calculation of switches in an automatic telephone system. In: Tore Olaus Engset: The man behind the formula. (1998)
9. Baynat, B., Eisenmann, P.: Towards an Erlang-Like formula for GPRS/EDGE network engineering. In: IEEE Int. Conf. on Communications (ICC). (June 2004)
10. Baynat, B., Boussetta, K., Eisenmann, P., Rached, N.B.: Towards an Erlang-Like formula for the performance evaluation of GPRS/EDGE networks with finite-length sessions. In: Proc. of 3rd IFIP-TC6 Networking Conference. (May 2004)
11. Berger, A., Kogan, Y.: Dimensioning bandwidth for elastic traffic in high-speed data networks. IEEE/ACM Transactions on Networking **8**(5) (October 2000) 643–654
12. Fredj, S.B., Bonald, T., Proutiere, A., Regnie, G., Roberts, J.: Statistical bandwidth sharing: A Study of congestion at flow level. In: Proc. of ACM Sigcomm. (August 2001)
13. Heyman, D., Lakshman, T., Neidhardt, A.: New method for analyzing feedback protocols with applications to engineering web traffic over the internet. In: Proc. of the ACM Sigmetrics. (June 1997)
14. Feldmann, A., Gilbert, A.C., Huang, P., Willinger, W.: Dynamics of IP traffic: A study of the role of variability and the impact of control. In: Computer Communication Review. (October 1999)
15. Ramadas, K., Jain, R.: WiMAX System Evaluation Methodology. Technical report, Wimax Forum (January 2007)