An Analytical Model for WiMAX Networks with Multiple Traffic Profiles and Throttling Policy

Sebastien Doirieux*, Bruno Baynat*, Masood Maqbool+, Marceau Coupechoux+
*Université Pierre et Marie Curie - Paris, FRANCE, {firstname.lastname}@lip6.fr
+TELECOM ParisTech & CNRS LTCI - Paris, FRANCE, {firstname.lastname}@telecom-paristech.fr

***Abstract*— In this paper, we present a simple and accurate analytical model for performance evaluation of WiMAX networks with multiple traffic profiles. This very promising access technology has been designed to support numerous kinds of applications having different traffic characteristics. One of the QoS parameters considered by the standard for traffic classes is the maximum sustained traffic rate (MSTR), which is an upper bound for user throughput. Taking into account MSTR implies the implementation of a throttling scheduling policy that regulates the user peak rate. Our models take into account this policy and provides closed-form expressions giving all the required performance parameters for each traffic profile at a click speed. The model is compared with extensive simulations that show its accuracy and robustness.**

I. INTRODUCTION

WiMAX (Worldwide Interoperability for Microwave Access) is a broadband wireless access technology which is 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 amendment IEEE 802.16e (mobile WiMAX) in 2005 [2]. A new standard, 802.16m, is currently under definition for providing even higher efficiency. On the other hand, the consortium WiMAX Forum was founded 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 certificate products (through inter-operability tests).

In order to accommodate various traffic types, different service categories have been introduced for WiMAX networks. For example, best effort (BE) is a service category that could handle web traffic. Each service category is characterized by its quality of service (QoS) parameters. One of the parameters associated with BE service is the maximum sustained traffic rate (MSTR). As defined in [2] (section 11.13.6), this is not the guaranteed rate but an upper bound. The procedure to implement this rate has been left open in the standard.

Taking into account this limited achievable user data rate is an important challenge while dimensioning a wireless network. Existing literature does not speak much about this subject. A detailed account of simulation based BE traffic performance evaluation in WiMAX networks can be found in [3], [4] and [5]. However, MSTR has not been considered in these papers. Mean information rate (MIR), a notion similar to MSTR, has been introduced in [6]. Authors have studied the performance of multi-profile internet traffic in presence of different MIR values for a WiMAX cell. They have used packet level simulations to evaluate the cell throughput performance for different number of users while considering possibility of multiple modulation schemes.

As far as analytical modeling is concerned, multi-class processor-sharing queues based models have been proposed in [7]–[9]. In these models, variability of radio channel conditions, an important property of wireless medium, has been taken into account. On the other hand, the idea of MSTR has not been touched in these articles. Earlier we have presented analytical models for mono-traffic [10], [11] and multi-traffic [12] BE without considering the parameter MSTR. Three generic scheduling policies: slot sharing fairness, throughput fairness and opportunistic scheduling were taken into account. In contrast to our existing work, models proposed in this paper are based on a fourth scheduling scheme, the throttling policy, in which maximum achievable user data rate is limited to MSTR. This in turn affects the resource utilization. The models proposed in this paper will offer the flexibility to network operators to dimension the WiMAX networks.

The organization of this paper is as follows. Modeling assumptions are presented in section II. The analytical models for mono/multi profile traffics are given in sections III and IV. Validation of models is presented in section V. At the end, section VI gives a conclusion of this work.

II. MODELING ASSUMPTIONS

In this section, we discuss the assumptions that have been considered in development of the model. Wherever required, related details of WiMAX system are also 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

1) We consider a single WiMAX cell and focus on the downlink part which is a critical portion of asymmetric data traffic.
2) We assume that amount of overhead in the TDD frame is fixed. 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.

3) The number of simultaneous mobiles that can be multiplexed in one TDD frame 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 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 mobile station (MS) using MCS_k is denoted m_k.

Channel assumption

4) The coding scheme used by a given mobile can change very often because of the high variability of the radio link quality. We assume that each mobile sends a feedback channel estimation on a frame by frame basis, and thus, the base station can change its coding scheme every frame. We thus associate a probability p_k with each coding scheme MCS_k, and assume that, at each time-step T_F, any mobile has a probability p_k to use MCS_k.

Traffic assumptions

5) We assume that there is a fixed number N of mobiles that are sharing the available bandwidth of the cell.
6) Each of the N mobiles 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). 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 periods, 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

8) At each frame, the scheduler tries to allocate the right number of slots to each active mobile in order to achieve its MSTR. If a mobile is in outage it does not receive any slot and its throughput is temporarily degraded. If at a given time the total number of available slots is not enough to satisfy the MSTR of all active users (not in outage), they all see their throughputs equally degraded.

III. Mono-traffic Analytical Model

1) Model description: A first attempt for modeling this system would be to develop a multidimensional 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 mobiles using coding scheme MCS_k, $0 \leq k \leq K$ (i.e., 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 mobiles, regardless of the coding scheme they use. The resulting CTMC is thus made of $N+1$ states as shown in Fig 1.

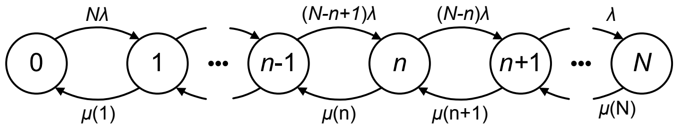


Fig. 1. General CTMC with state-dependent departure rates.

- A transition out of a generic state n to a state $n+1$ occurs when a mobile in OFF period starts its transfer. This “arrival” transition corresponds to one mobile 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:
$$\lambda = \frac{1}{\bar{t}_{off}}. \tag{1}$$
- A transition out of a generic state n to a state $n-1$ occurs when a mobile 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 mobiles are active.

Obviously, the main difficulty of the model resides in estimating the aggregate departure rates $\mu(n)$ that strongly depend on the chosen scheduling policy. Focusing in this paper on the throttling policy, we now explain how to do so when considering this particular policy. Note that our previous works [10]–[12] have considered “full-capacity” policies (slot sharing fairness, throughput fairness and opportunistic scheduling) and derived for them very different expressions of the $\mu(n)$.

2) Departure rates: In order to estimate the average departure rates $\mu(n)$, we first define the following quantities.

To compensate losses due to outage, we consider a slightly greater instantaneous bitrate than the MSTR, the Delivered BitRate:
$$DBR = \frac{MSTR}{1-p_0}. \tag{2}$$
A mobile using MCS_k needs a mean number of $\bar{g}_k$ slots per frame to reach its DBR:
$$\bar{g}_k = \frac{DBR\ T_F}{m_k}. \tag{3}$$
Obviously, no slot is allocated to a mobile in outage so $\bar{g}_0 = 0$.

From this, we then deduce $\bar{g}$, the average number of slots per frame needed by a mobile to obtain its MSTR:

$$\bar{g} = \sum_{k=1}^{K} p_k \bar{g}_k. \tag{4}$$

Knowing $\bar{g}$, we can now express the aggregate departure rates $\mu(n)$ as follows:

$$\mu(n) = \frac{N_S}{max\,(n\bar{g}, N_S)} n \frac{MSTR}{\bar{x}_{on}}. \tag{5}$$

The last part of this expression ($n\frac{MSTR}{\bar{x}_{on}}$) corresponds to the rate at which any of the n active mobiles completes its transfer, assuming that there are always enough available slots in the frames to satisfy the MSTR. The first part of this expression ($\frac{N_S}{max(n\bar{g},N_S)}$) represents the ratio of the global departure rate achieved by the n concurrent active transfers. Indeed, when there are n active mobiles, they need $n\bar{g}$ slots in average to obtain their MSTR.

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

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

whereρ is given by:

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

and $\pi(0)$ is obtained by normalization.

The performance parameters of this system can be derived from the steady-state probabilities as follows.

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

$$\bar{Q} = \sum_{n=1}^{N} n\, \pi(n). \tag{8}$$

The average number $\bar{D}$ of departures (i.e., mobiles completing their transfer) per unit of time is given by:

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

From Little's law, we can derive the average duration $\bar{t}_{on}$ of an ON period (duration of an active transfer):

$$\bar{t}_{on} = \frac{\bar{Q}}{\bar{D}}, \tag{10}$$

and compute the average throughput $\bar{X}$ obtained by each mobile in active transfer as:

$$\bar{X} = \frac{\bar{x}_{on}}{\bar{t}_{on}} = \frac{\bar{x}_{on} \sum_{n=1}^{N} \mu(n)\, \pi(n)}{\sum_{n=1}^{N} n\, \pi(n)}. \tag{11}$$

Finally, we can express the average utilization $\bar{U}$ of the TDD frame, as a weighted sum of the ratios between the mean number of slots needed by the n mobiles to reach their MSTR and the mean number of slots they obtain:

$$\bar{U} = \sum_{n=1}^{N} \frac{n\bar{g}}{max\,(n\bar{g}, N_S)} \pi(n). \tag{12}$$

Lastly, note that when $max\,(N\bar{g}, N_S) = N_S$, i.e., when the resources of the system are sufficient to satisfy the $MSTR$ even if all the N mobiles of the cell are in active transfer, we can easily demonstrate that the average throughput of each active mobile (obtained from relation 11) becomes $\bar{X} = MSTR$.

IV. Multi-traffic Extension

We now consider that users are divided into R classes of traffic with specific traffic profiles $(MSTR_r, \bar{x}^r_{on}, \bar{t}^r_{off})$. Each mobiles of a given class r thus has a maximum instantaneous throughput $MSTR_r$ and generates an infinite-length ON/OFF traffic, with an average ON size of $\bar{x}^r_{on}$ bits and an average OFF duration of $\bar{t}^r_{off}$ seconds.

We assume that there is a fixed number N_r of mobiles belonging to each class in the cell. Finally, mobiles of different classes may have different channel models. A mobile of class r thus has a probability p_{kr} of using MCS_k.

We saw in the mono-trafic model, when expressing the steady-state probabilities $\pi(n)$, that these probabilities as well as all performance parameters mostly depend on the traffic profile $(MSTR, \bar{x}_{on}, \bar{t}_{off})$ through a single aggregated parameter ρ given by relation 7. The key assumption of this multi-traffic extension is to suppose that all the performance parameters of the resulting multi-class model are still dependent of the traffic profiles through a set of aggregated parameters ρ_r given by:

$$\rho_r = \frac{\bar{x}^r_{on}}{\bar{t}^r_{off} MSTR_r}. \tag{13}$$

As a consequence, we can transform any class-r profile $(MSTR_r, \bar{x}^r_{on}, \bar{t}^r_{off})$ into an equivalent profile $(MSTR, \bar{x}_{on}, \bar{t'}^r_{off})$, such that:

$$\frac{\bar{x}_{on}}{\bar{t'}^r_{off} MSTR} = \frac{\bar{x}^r_{on}}{\bar{t}^r_{off} MSTR_r}. \tag{14}$$

By doing so for all classes, we transform the original system into an equivalent system where all classes of traffic have the same maximum instantaneous throughput $MSTR$, the same average ON size $\bar{x}_{on}$, and different average OFF durations $\bar{t'}^r_{off}$.

With this transformation, the equivalent system can be described as a multi-class closed queuing network with two stations (see Fig. 2):

1) An IS (infinite-server) station that models mobiles in OFF periods. This station has class-dependent service rates λ_r given by:

$$\lambda_r = \frac{1}{\bar{t'}^r_{off}}; \tag{15}$$

2) A PS (processor sharing) station that models active mobiles. This station has class-independent service rates

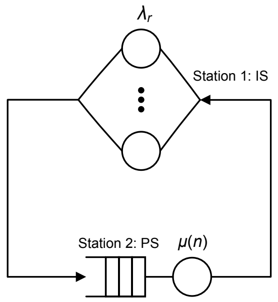


Fig. 2. Closed-queueing network.

$\mu(n)$ that in turn depend on the total number active mobiles (whatever their classes).

However, we cannot directly use the same expression of the average departure $\mu(n)$ obtained in the mono-traffic case (Eq. 5). Indeed, if we look at the expression of the steady-state probabilities derived for the mono-traffic model (Eq. 6), we can see that they not only depend on the traffic profile through the aggregated parameter ρ, but also through the parameter $\bar{g}$ that represents the average number of slots per frame needed by a mobile to obtain its $MSTR$. We thus propose to use a very similar expression for $\mu(n)$:

$$\mu(n) = \frac{N_S}{max\,(\bar{g}(n), N_S)} n \frac{MSTR}{\bar{x}_{on}}, \tag{16}$$

in which $MSTR$ and $\bar{x}_{on}$ are the common values of the equivalent multi-class profiles, and $\bar{g}(n)$ is the average number of slots per frame needed by n mobiles to obtain their maximum throughput.

In order to derive an expression for $\bar{g}(n)$ that takes into account the different classes of traffic, we first express DBR_r, the actual bitrate needed by a mobile of class r in order to reach its $MSTR_r$ (while compensating losses due to outage):

$$DBR_r = \frac{MSTR_r}{1 - p_{0r}}. \tag{17}$$

We then define $\bar{g}_r$, the mean number of slots needed by a mobile of class r to obtain its $MSTR_r$, as:

$$\bar{g}_r = \sum_{k=1}^{K} p_{kr} \frac{DBR_r\ T_F}{m_k}. \tag{18}$$

Secondly, we estimate the probabilities $\alpha_r(n)$ that an active mobile belong to class r knowing that n mobiles are active (i.e., n customers are in the PS station). These probabilities are obvious when $n = N$:

$$\alpha_r(N) = \frac{N_r}{N}, \tag{19}$$

and closely approximated when $n = 1$ by:

$$\alpha_r(1) = \frac{N_r \rho_r}{\sum_{i=1}^{R} N_i \rho_i}. \tag{20}$$

Knowing $\alpha_r(1)$ and $\alpha_r(N)$, we then suppose that the $\alpha_r(n)$ are a linear function of n:

$$\alpha_r(n) = an + b, \tag{21}$$

with

$$a = \frac{\alpha_r(N) - \alpha_r(1)}{N - 1}, \tag{22}$$

and

$$b = \frac{N\alpha_r(1) - \alpha_r(N)}{N - 1}. \tag{23}$$

Lastly, we express the average parameter $\bar{g}(n)$ as:

$$\bar{g}(n) = \sum_{r=1}^{R} n\, \alpha_r(n) \bar{g}_r. \tag{24}$$

Note that the $\alpha_r(n)$ probabilities can alternately be obtained by considering a multi-dimensional Markov chain which states $(n_1, ..., n_R)$ correspond to the detailed distribution of the current active mobiles of each class in the system. From the numerical resolution of this chain we can derive the exact values of the $\alpha_r(n)$ probabilities. We have checked on several examples that the exact $\alpha_r(n)$ probabilities are very well estimated by the linear approximation we propose above. In addition, the impact of this approximation is very limited as it only matters states n such that $n\bar{g}(n) < N_S$ (see Eq. 16). Finally it is important to emphasize that the use of this approximation enables to avoid the exponential complexity of solving a multi-dimensional Markov chain.

A direct extension of the BCMP theorem [13] for stations with state-dependent rates can now be applied to this closed queueing network. The population vector is denoted by $\overrightarrow{N} = (N_1, ..., N_R)$. The detailed steady-state probabilities can then be expressed as follows:

$$\pi(\overrightarrow{n}) = \pi(\overrightarrow{n_1}, \overrightarrow{n_2}) = \frac{1}{G} f_1(\overrightarrow{n_1}) f_2(\overrightarrow{n_2}), \tag{25}$$

where $\overrightarrow{n_i} = (n_{i1}, ..., n_{iR})$, n_{ir} being the number of class-r mobiles present in station i,

$$f_1(\overrightarrow{n_1}) = \frac{1}{n_{11}! ... n_{1R}!} \frac{1}{(\lambda_1)^{n_{11}} ... (\lambda_R)^{n_{1R}}} \tag{26}$$

and

$$f_2(\overrightarrow{n_2}) = \frac{(n_{21} + ... + n_{2R})!}{n_{21}! ... n_{2R}!} \frac{1}{\prod_{k=1}^{n_2} \mu(k)}, \tag{27}$$

and G is a normalization constant:

$$G = \sum_{\overrightarrow{n_1} + \overrightarrow{n_2} = \overrightarrow{N}} f_1(\overrightarrow{n_1}) f_2(\overrightarrow{n_2}). \tag{28}$$

All the performance parameters of interest can be derived from the steady-state probabilities as follows. The average number of customers of class r in station 2, i.e., the average number of class-r active mobiles, denoted by $\bar{Q}_r$, is given by:

$$\bar{Q}_r = \sum_{\overrightarrow{n_1} + \overrightarrow{n_2} = \overrightarrow{N}} n_{2r}\, \pi(\overrightarrow{n_1}, \overrightarrow{n_2}). \tag{29}$$

Let $\bar{D}_r$ be the average number of class-r customers departing from station 2 by unit of time, i.e., the average number of class-r mobiles completing their download by unit of time.

$\bar{D}_r$ can be expressed as:

$$\bar{D}_r = \sum_{\overrightarrow{n_1}+\overrightarrow{n_2}=\overrightarrow{N}} \mu_r(\overrightarrow{n_2})\,\pi(\overrightarrow{n_1},\overrightarrow{n_2}), \tag{30}$$

where $\mu_r(\overrightarrow{n_2})$ is the departure rate of class-r mobiles when there are $\overrightarrow{n_2}$ active mobiles:

$$\mu_r(\overrightarrow{n_2}) = \frac{N_S}{max\left(\bar{g}(\overrightarrow{n_2}), N_S\right)} n_{2r} \frac{MSTR_r}{\bar{x}^r_{on}}, \tag{31}$$

with

$$\bar{g}(\overrightarrow{n_2}) = \sum_{r=1}^{R} n_{2r}\bar{g}_r. \tag{32}$$

The average download duration of class-r mobiles, $\bar{t}^r_{on}$, is nothing but the average sojourn time of class-r customers in station 2, and is obtained from Little law:

$$\bar{t}^r_{on} = \frac{\bar{Q}_r}{\bar{D}_r}. \tag{33}$$

We can then calculate the average throughput $\bar{X}_r$ obtained by customers of class r during their transfer as:

$$\bar{X}_r = \frac{\bar{x}^r_{on}}{\bar{t}^r_{on}} = \frac{\bar{x}^r_{on} \displaystyle\sum_{\overrightarrow{n_1}+\overrightarrow{n_2}=\overrightarrow{N}} \mu_r(\overrightarrow{n_2})\,\pi(\overrightarrow{n_1},\overrightarrow{n_2})}{\displaystyle\sum_{\overrightarrow{n_1}+\overrightarrow{n_2}=\overrightarrow{N}} n_{2r}\,\pi(\overrightarrow{n_1},\overrightarrow{n_2})}. \tag{34}$$

Finally, we can express the utilization $\bar{U}$ of the TDD frame as:

$$\bar{U} = \sum_{\overrightarrow{n_1}+\overrightarrow{n_2}=\overrightarrow{N}} \frac{\bar{g}(\overrightarrow{n_2})}{max\left(\bar{g}(\overrightarrow{n_2}), N_S\right)} \pi(\overrightarrow{n_1},\overrightarrow{n_2}). \tag{35}$$

V. Validation through Simulations

In this section we discuss the validation of our analytical models through extensive simulations. We also study the effect of more complex traffic and channel models. We start with details of simulator followed by results for mono/multi-traffic models.

A. Simulation Details

1) System Parameters: System bandwidth is assumed to be 10 MHz. The downlink/uplink ratio of the WiMAX TDD frame is considered to be approximately 2/3 (32 OFDM symbols on the downlink and 15 on the uplink). We assume for the sake of simplicity that the protocol overhead is of fixed length (2 symbols) although in reality it is a function of the number of scheduled users. Considering subcarrier permutation PUSC on the downlink, there are 30 sub-channels and a slot is defined as 2 OFDM symbols and one subchannel. With above parameters, the total number of data slots (excluding overhead) per TDD downlink sub-frame is $N_S = 30(32-2)/2 = 450$.

2) Traffic Parameters: Values of MSTR, mean ON data volume (main page and embedded objects) and OFF period (reading time) for both mono and multi-traffic are given in Tab. I.

In the validation study, we assume that the ON sizes are 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 robustness study. Recall that the mean value of the truncated Pareto distribution is given by:

$$\bar{x}_{on} = \frac{\alpha b}{\alpha - 1}\left[1 - (b/q)^{\alpha-1}\right], \tag{36}$$

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. These have been taken as hundred times and thousand times the mean value respectively. The mean value in both cases (higher and 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]. The corresponding values of parameter b for higher and lower cutoff are calculated using Eq. 36. Pareto parameters, used in simulations for $\bar{x}_{on} = 3$ Mbps, are summarized in Tab. II.

3) Channel Models: A generic method for describing the channel between the BS (Base Station) 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 $P_T = (p_{ij})_{0\le i,j\le K}$. Note that an additional state (state 0) is introduced to take into account outage (SINR is below the minimum radio quality threshold). 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 $(p_i)_{0\le i,j\le K}$. 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.

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 the following equation:

$$P_T(a) = aI + (1-a)P_T(0) \quad 0 \le a \le 1,$$

where I is the identity matrix and parameter a is a measure of the channel memory. A MS indeed 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

TABLE I
TRAFFIC PARAMETERS.

Parameter	Mono-traffic	Multi-traffic	
		Class 1	Class 2
$MSTR$ [Kbps]	512 & 2048	1024	2048
$\bar{x}_{on}$ [Mbps]	3	3	3
$\bar{t}_{off}$ [s]	3	3	6

TABLE II
PARETO PARAMETERS FOR $\bar{x}_{on} = 3$ MBPS.

Parameter	Value
Shape parameter α	1.2
Lower cutoff q	300 Mbits
Higher cutoff q	3000 Mbits
b for lower cutoff	712926 bits
b for higher cutoff	611822 bits

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. For the sake of comparison, the overall MCS probabilities in the combined model are the same as those of the memoryless and average models.

Three models are thus considered: the memoryless, the average, and the combined channel models. Corresponding stationary probabilities are given in Tab. III. Stationary probabilities for the combined model are obtained by averaging corresponding values of good and bad model stationary probabilities. Considered MCS and for each of them, the number of bits transmitted per slot are given in Tab. IV.

B. Simulation Results

In this section, we first present the validation/robustness study results for mono-traffic model and then validation study results for multi-traffic model. The output parameters in consideration are $\bar{U}$, $\bar{X}$, $\bar{Q}$, and $\pi(n)$ (see Sections III and IV).

1) Mono-Traffic: To validate the mono-traffic model, 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. Distributions of ON size and OFF period are exponential and the memoryless channel model is considered.

Results from performance validation are shown in Fig. 3. For different number of users in the cell, the values of analytical model and simulation differ only by 2% at most.

We now move to the robustness study, where assumptions concerning traffic and channel models made by the analysis are relaxed in simulations. The results for this analysis are shown in Fig. 4. As can be noticed on Fig. 4(a), our model works well for different values of MSTR.

TABLE III
STATIONARY PROBABILITIES FOR THREE CHANNEL MODELS.

Channel model	Memoryless	Average	Combined	
			50% MS good	50% MS bad
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 IV
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	$m_4 = 144$

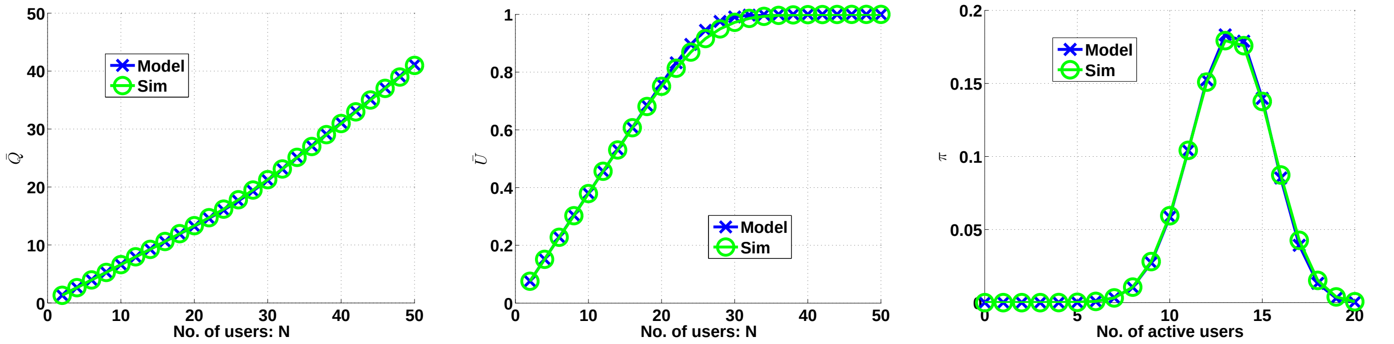
It is also clear (see Fig. 4(b)) that considering a truncated Pareto distribution has little influence on the design parameters. The average relative error between analytical results and simulations stays below 10% for all sets. This is mainly due to the fact that the distribution is truncated and is thus not heavy tailed. But even with a high cutoff value, the exponential distribution provides a very good approximation.

Similarly, the results illustrated on Fig. 4(c) also prove 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.

2) Multi-Traffic: In multi-traffic scenario, we consider two different classes of traffic. Each class is characterized by particular values of MSTR, $\bar{x}_{on}$ and $\bar{t}_{off}$ (see Tab. I). Fig. 5 shows that simulation and analysis provide similar results not only for the overall system performance but also for each class (maximum difference is below 6%). As expected, users obtain their respective MSTR at low load and when load increases, they see their throughput proportionally degraded (Fig. 5(c)).

VI. CONCLUSION

As deployment of WiMAX networks is underway, need arises for operators and manufacturers to develop dimensioning tools. In this paper, we have presented an analytical model for WiMAX networks taking into account multiple data traffic profiles and the QoS parameter MSTR defined by the standard. Our model, based on a product-form closed queueing network, is able to instantaneously provide Erlang-like performance parameters such as throughput per user for each profile or channel utilization. Therefore it will render possible efficient and advanced dimensioning studies. Moreover, the simple

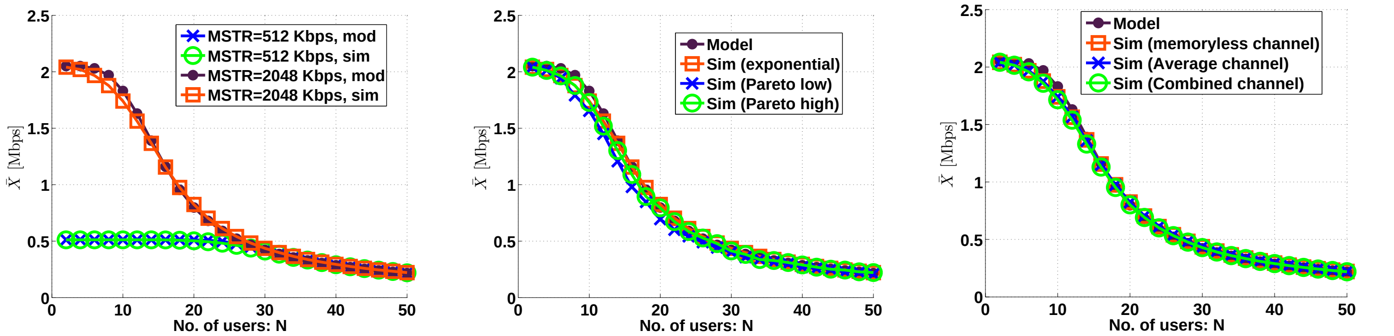


(a) Average number of active users.

(b) Average utilization.

(c) Stationary probabilities for $N = 20$.

Fig. 3. Validation study with MSTR $= 512$ Kbps, $\bar{x}_{on} = 3$ Mbits and $\bar{t}_{off} = 3$ s.

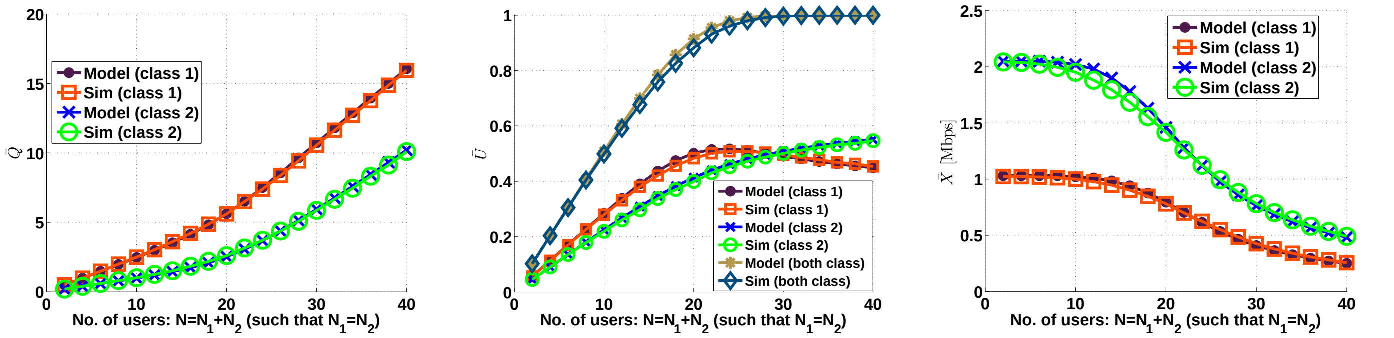


(a) Average throughput per user for different values of MSTR (512 and 2048 Kbps).

(b) Average throughput per user for MSTR $=$ 2048 Kbps and different traffic distributions.

(c) Average throughput per user for MSTR $=$ 2048 Kbps and different channel models.

Fig. 4. Robustness study with $\bar{x}_{on} = 3$ Mbits and $\bar{t}_{off} = 3$ s.



(a) Average number of active users.

(b) Average utilization.

(c) Average throughput per user.

Fig. 5. Performance validation for multi-traffic with class 1 parameters: MSTR $= 1024$ Kbps, $\bar{x}_{on} = 3$ Mbits and $\bar{t}_{off} = 3$ s and class 2 parameters: MSTR $= 2048$ Kbps, $\bar{x}_{on} = 3$ Mbits and $\bar{t}_{off} = 6$ s.

nature of our model makes it flexible to be customized to scenario specific requirements. Extensive simulations have validated the model's assumptions and showed its robustness towards more complex traffic distributions and channel models.

References

[1] "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems," 2004.

[2] "Draft IEEE std 802.16e/D9. IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems. Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands," 2005.

[3] S. Kim and I. Yeom, "Performance analysis of best effort traffic in ieee 802.16 networks," in *IEEE TRANSACTIONS ON MOBILE COMPUTING*, 2008.

[4] G. Leonardi, A. Bazzi, G. Pasolini, and O. Andrisano, "Ieee802.16e best effort performance investigation," in *Proc. IEEE ICC'07*, June 2007.

[5] F. Hou, J. She, P.-H. Ho, and X. Shen, "Performance analysis of weighted proportional fairness scheduling in ieee 802.16 networks," in *Proc. IEEE ICC'08*, May 2008.

[6] D. Sivchenko, N. Bayer, B. Xu, V. Rakocevic, and J. Habermann,

"Internet Traffic Performance in IEEE 802.16 Networks," in *Proc. of 12th European Wireless Conference*, April 2006.

[7] S. Borst, "User-Level Performance of Channel-Aware Scheduling Algorithms in Wireless Data Networks," in *IEEE Infocom*, 2003.

[8] T. Bonald and A. Proutiere, "Wireless downlink channels: User performance and cell dimensioning," in *ACM Mobicom*, 2003.

[9] S. Liu and J. Virtamo, "Performance Analysis of Wireless Data Systems with a Finite Population of Mobile Users," in *19th International Teletraffic Congress*, 2005.

[10] G. Nogueira, B. Baynat, M. Maqbool, and M. Coupechoux, *Book chapter in "WiMAX Networks Planning and Optimization"*, First ed. Auerbach Publications, CRC Press, Taylor & Francis Group, 2008.

[11] B. Baynat, G. Nogueira, M. Maqbool, and M. Coupechoux, "An efficient analytical model for the dimensioning of wimax networks," in *Proc. of 8th IFIP-TC6 Networking Conference*, May 2009.

[12] B. Baynat, S. Doirieux, G. Nogueira, M. Maqbool, and M. Coupechoux, "An efficient analytical model for wimax networks with multiple traffic profiles," in *Proc. of ACM/IET/ICST IWPAWN*, September 2008.

[13] F. Baskett, K. Chandy, R. Muntz, and F. Palacios, "Open, closed, and mixed networks of queues with different classes of customers," in *Journal of the Association of Computing Machinery*, April 1975, pp. 22(2): 248–260.

[14] A. Feldmann, A. C. Gilbert, P. Huang, and W. Willinger, " Dynamics of IP traffic: A study of the role of variability and the impact of control," in *Computer Communication Review, a publication of ACM SIGCOMM*, October 1999.

[15] K. Ramadas and R. Jain, "WiMAX System Evaluation Methodology," Wimax Forum, Tech. Rep., January 2007.