Wireless Netw (2011) 17:629–644
DOI 10.1007/s11276-010-0302-y

A hybrid CAC algorithm for maximizing downlink capacity of M-WiMAX systems

Georgios Theodoridis · Fotini-Niovi Pavlidou

Published online: 28 November 2010

Abstract Due to the implementation of an Adaptive Modulation-Coding (AMC) mechanism in the 802.16e physical layer, each connection's bandwidth requirements cannot be statically computed, but they derive as a function of the terminals' instantaneous Signal-to-Noise Ratio (SNR). Therefore, the Connection Admission Control (CAC) mechanism fails to establish an efficient policy that would optimally exploit the system's resources. In this respect, the present paper formulates initially an adequate method for statistically calculating the average capacity of a Mobile-WiMAX system, according to the distribution of the several attenuation factors that affect the signal along the propagation path. In parallel, based on the processing of the terminal's SNR samples, a sophisticated algorithm for filtering out any misleading measurements is developed. This second method aims at predicting the upcoming Modulation-Coding state of every connection, so as to acquire a short-term view of its spectrum demands. Finally, all the above info collected from both the estimation methods is utilized by the CAC procedure in order to perform an as accurate as possible computation of the resource availability. As a result, this hybrid approach succeeds in maximizing the network revenue through significantly increasing the number of concurrently serviced connections while guaranteeing their Quality of Service (QoS) standards.

G. Theodoridis (✉) · F.-N. Pavlidou
Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
e-mail: getheod@auth.gr

F.-N. Pavlidou
e-mail: niovi@auth.gr

Keywords 802.16e · Admission control · Capacity estimation · M-WiMAX · Resource allocation

1 Introduction

Although the broadband era has been so far boosted by fibre optics and DSL (Digital Subscriber Line), the development of a reliable and cost-efficient wireless and eventually mobile access system that would expand high-speed connectivity beyond the limitations of the wired infrastructure, has shortly become a requisite of primary importance. Under these circumstances, following the ultimate success of IEEE 802.11, the IEEE 802.16 standard, widely known under the term WiMAX (Worldwide Interoperability for Microwave Access), was introduced [1] as the prevailing solution for providing Broadband Wireless Access (BWA) over a broader coverage area, in a Metropolitan Area Network (MAN) manner. In addition, since IEEE 802.16 was initially designed solely for fixed terminals, in 2005, the 802.16e amendment was presented, incorporating all the adequate changes and enhancements that are necessary for supporting mobility [2].

According to its specifications, M(obile)-WiMAX is expected to achieve downlink datarates of 70 Mbps for a distance that shall range up to 15 kms. In order to maintain such a high performance against the harsh mobile conditions, its physical layer combines a set of the most innovative technologies, such as OFDMA (Orthogonal Frequency Division Multiple Access), Adaptive Modulation-Coding (AMC), power control and MIMO (Multiple Input/Output) techniques [3, 4]. Common feature of all the aforementioned mechanisms is that they are particularly chosen so as to guarantee adaptability for the whole range of potential propagation environments, via the dynamic

configuration of the signal transmission process. At the same time, in contrast to the detailed description of the physical and MAC (Medium Access Control) layers, the exact protocols implemented at the IEEE 802.16e upper layers are not strictly defined by the standard, but they have been left open to be decided on per vendor basis, so as to facilitate the maximum possible flexibility of the overall system. In consequence, the optimization of the M-WiMAX Radio Resource Management (RRM) routines under these disparate conditions, is regarded as a top priority, yet challenging, task.

Specifically, one of the main complexity issues rises from the fact that, due to the AMC function, the Modulation-Coding Scheme (MCS) of each connection is dictated on per timeframe basis by the perceived SNR. Thus, the number of forwarded data bits per transmission symbol derives as a function of the instantaneous channel conditions encountered by the corresponding Mobile Station (MS). Hence, given the inherently stochastic attributes of the mobile environment, the connections' bandwidth requirements present intense fluctuations even for the case of data flows with constant bitrate. As a result, the efficiency of the Connection Admission Control (CAC) process is severely degraded, since the future capacity of the M-WiMAX network fails to be a priory approximated. In particular, an excessively pessimistic CAC approach, i.e. over-reservation of resources, shall lead to the unjustifiable blockage of new connections, while on the other hand, the underestimation of the flows' spectrum requirements shall eventually cause the violation of their QoS (Quality of Service) standards because of shortage of resources. In consequence, it becomes evident that the CAC must include an adequate mechanism for overcoming the variations that are introduced by the AMC operation, so as to be able to form a solid estimation of the network's needs.

In this framework, the present paper firstly introduces an accurate method for analytically calculating the mean value of the total bandwidth demands for a given set of flows, based on the MSs' prospective spatial distribution as well as the Line-of-Sight (LOS) probability model. In parallel, the feedback of the MSs' SNR (Signal-to-Noise Ratio) is exploited through a heuristic sampling and averaging procedure, in order the bandwidth requirements for the forthcoming narrow time-window to be precisely computed. Finally, the above statistical and measurement-based results, which respectively provide a long and short-term estimation of the system's capacity, are integrated into a single hybrid CAC scheme. This way, CAC's fidelity is radically improved, as the admittance decision is based on the most solid and overall knowledge of the network's status.

The rest of the paper is organized as follows: Sect. 2 summarizes a state-of-the-art review of all the related work and Sect. 3 briefly presents the specific 802.16e technologies that mainly affect RRM procedure. In Sect. 4, the herein introduced CAC algorithm is thoroughly described, covering its dual functionality, while, in Sect. 5, the prediction's accuracy and the overall algorithm's performance is evaluated via an adequate simulation platform. At the end, Sect. 6 draws the final conclusions.

Before continuing, it must be clarified that the proposed scheme is also applicable to the M-WiMAX uplink sub-system. Nevertheless, since the exact implementation differs according to the distinctive elements of each sub-system, for reasons of coherence, in this paper we shall solely study the downlink path.

2 Related work

The realization of the high expectations stemming from the new standard calls for an optimum RRM approach against the scarcity of radio resources and the augmented traffic intensity. Therefore, the issue of maximizing CAC's effectiveness within the WiMAX architecture has been so far the area of a notable amount of research activity, towards the common target of increasing network's performance in terms of fairness, reliability and revenue. To achieve these goals, the available studies either extend well-established methodologies in order to suite to the particular features of the WiMAX configuration or develop novel tactics specially tailored for this purpose.

Being more specific, in Niyato et al. [5] utilize a game theoretic technique so as to share the common channel among the various connections in a fair way from both the network's and the subscribers' point of view. The same authors, in Niyato and Hossain [6], implement a CAC mechanism according to which a new connection is accepted with a dynamically adjustable probability threshold, depending on the length of the packets' queue and the radio link quality; the number of concurrent connections is treated as a Markov chain and the transition matrix is calculated. Moreover, in Rong et al. [7], introduce Adaptive Power Allocation (APA) as the regulating tool, in order to achieve balance between the network revenue and the subscribers' QoS; CAC gathers the available info from the APA routine, returning also all the necessary feedback.

Especially for the M-WiMAX case, the authors of [8] compute the effect of the physical and MAC configuration upon the overall system's efficiency along with the specific impact of AMC on the holding time of elastic calls in presence of mobility. Furthermore, studying the AMC's effect from the reverse perspective, the authors of [9] augment the aggregate capacity via limiting the transmission power per MS. The AMC level of each MS is reduced,

causing an increment in subchannel requirement; on the other hand, the corresponding decrement in transmit power minimizes inter-cell interference, causing the maximization of SINR (Signal to Interference and Noise Ratio). Alternatively, as means of overcoming the stochastic nature of RRM in IEEE 802.16e, Lee et al. [10] base their predictions on the Central Limit Theorem, according to which, the sum of a large number of random variables (each connection's bandwidth requirements) can be rather accurately approximated by a single variable following normal distribution. Furthermore, from the QoS point of view, the CAC algorithm in Tsang et al. [11] allows the provision of different QoS guarantees to each one of the available traffic classes through a threshold-based scheme, while the CAC mechanism in Kalikivayi et al. [12] is built upon the bandwidth and delay restrictions imposed by each flow.

In a parallel approach, in mobile scenarios, the efficiency of the resource reservation policies can be significantly increased by the a-priori knowledge of the MSs' mobility pattern. Although most of the researches in this field refer to the manipulation of the handover phenomenon, their applicability can be also extended to the CAC mechanisms of wireless networks with dynamic physical configuration (e.g. M-WiMAX), as the alterations of the MSs' link quality is directly connected with the path across which they travel. In particular, in Zonoozi and Dassanayake [13], a mathematical formulation is developed for systematic tracking of the random movement of a mobile station in a cellular environment, incorporating a large variety of mobility parameters. Moreover, in order to pinpoint the MSs' upcoming locations, the authors of [14] suggest the submission of periodical location and velocity reports from the terminals to the NCC (Network Control Centre), while in Soh and Kim [15], Soh et al. are making virtue of the mobile positioning technology presuming the widespread availability of digital road maps. On the contrary, in Samaan and Karmouch [16] and Ma et al. [17], no assumptions are made regarding the availability of information about the movements' history, but each terminal's trajectory is predicted by using a personalized user profile (preferences, goals and analysed spatial information).

3 M-WiMAX: pivotal technologies

3.1 Physical layer

Multiple access to the IEEE 802.16e common radio channel is carried out upon OFDMA technology. Along the *frequency* axis, the available bandwidth, W_{tot}, is divided into numerous (C_{tot}) orthogonal and overlapping subcarriers. From these C_{tot} subcarriers only a number of C_{dat} subcarriers is used for the transmission of the users' data, while the rest are reserved for signalling purposes. In parallel, for higher resource granularity, along the *time* axis, the timeline is divided into sequential frames of T_{frm} duration which are further split into L_{tot} timeslots, in a TDMA manner. Furthermore, the two-dimension combination of C_{sch} subcarriers and L_{sch} timeslots forms the minimum bandwidth allocation unit, i.e. one subchannel (Fig. 1). The grouping of subcarriers into a subchannel is called permutation and it can be carried out in two different modes, according to which one subchannel can cover either distributed (diversity permutation) or adjacent (contiguous permutation) subcarriers [4]. Finally, since each timeslot holds a single modulation symbol, the overall pool of resources to be managed at the CAC level is equal to the aggregate number of data-symbols (timeslots) per frame, i.e $C_{\text{dat}} \cdot L_{\text{tot}}$. Thus, the BS's maximum symbol-rate, in symbols/s:

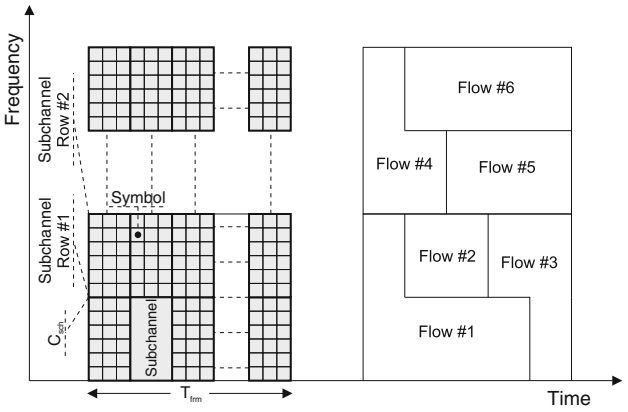


Fig. 1 General structure of a OFDMA frame

$$\text{SR}_{\max} = \frac{C_{\text{dat}} \cdot L_{\text{tot}}}{T_{\text{frm}}} \tag{1}$$

Despite the frequency overlap, the subcarriers' orthogonality allows the error-free demultiplexing of the simultaneous signals via a RFFT (Reverse Fast Fourier Transformation) procedure. Hence, besides the obvious advantage of augmenting (even doubling) the amount of exploitable bandwidth, the utilization of OFDMA also speeds up the transponders' reconfigurability, since no zone-pass filters are any longer necessary and all demultiplexing processes are carried out via execution of adequate software. Moreover, as each subcarrier occupies a vary narrow frequency band, it can be safely considered to be subject to flat fading [18].

Apart from OFDMA, one of the key characteristics of M-WiMAX physical layer, is the utilization of an AMC mechanism, which allows the system to maintain high performance under the unpredictable and complex mobile conditions. Specifically, depending on the transmission's

SNR level, the BS (Base Station)—MS communication automatically adopts one the seven available MCSs (Modulation-Coding Schemes) that are predefined by the standard. As a result, if $B(e)$ is the number of data-bits per symbol (except for coding overhead) that are forwarded under the eth MCS (MC $= e$), then, in order to sustain a datarate of DR bits/s, the symbol-rate must be equal to

$$\mathrm{SR} = \frac{\mathrm{DR}}{B(e)} \tag{2}$$

As it becomes evident from Eq. 2, the system's capacity cannot be statically foreseen, since, for given datarate demands, the necessary resource allocation to each MS derives as a function of its SNR. The set, E, of available MCSs are summarised in Table 1 along with the corresponding SNR thresholds ($\mathrm{SN}_{\mathrm{th},e}$) and data-bits per symbol ($B(e)$), $\forall e \in \mathbf{E} \equiv \{1, \ldots, E\}, E = 7$ [19]. It must also be noted that if SNR falls below $\mathrm{SN}_{\mathrm{th},1}$ (MC $= 0$), no transmission takes place.

3.2 QoS procedures

All modern networks have to meet the challenge of consistently servicing a great variety of applications with disparate traffic descriptors (nominal datarate, delay, jitter) in a way that must be transparent to the end-user. To accomplish this mission, IEEE 802.16e performs a categorization of the data flows into five QoS classes in order them to be forwarded within a priority-based scheme [18].

- UGS—Unsolicited Grant Service. The amount of resources reserved for each connection is constant and non-negotiable. The sustainable datarate of the connection is declared from the MS to the BS in advance and the necessary number of subchannels per timeframe is statically allocated. It is appropriate for providing services in a circuit-switched manner, e.g. uncompressed voice, similarly to the CBR (Constant Bit Rate) class of the ATM (Asynchronous Transfer Mode) architecture.
- rtPS—Real Time Polling Service. For a rtPS flow three parameters are defined: (a) Maximum Sustainable Traffic Rate (MSTR), (b) Minimum Reserved Traffic Rate (MRTR) and c) Maximum Latency. A rtPS connection is pre-allocated resources equal to MRTR, while any bursts are forwarded within the MSTR boundaries under the latency limitations. The BS polls the MS at specific time intervals based on the QoS requirements of the data-flow, so as the MS to transmit only after the clearance of the BS. In spite of the signalling overhead, the dynamic allocation guarantees the more efficient management of the spectrum. This class is suitable for transmitting compressed multimedia.
- ErtPS—Extended real-time Polling Service. In order to minimize the signalling overhead and delay imposed in rtPS, no polling is carried out. On the contrary, the clearance procedure is a-priori scheduled, according to the QoS limitations that are declared at connection setup time. This way, ErtPS is especially designated for supporting real-time applications that generate periodical, variable sized data packets, such as VoIP.
- nrtPS—Non Real Time Polling Service. It is similar to the rtPS case, with the only difference that the BS polling time intervals are not so rigid and the MS can also use contention request opportunities. Therefore, since no datarate or latency guarantees are provided, nrtPS is more adequate for TCP-based connections, i.e. the vast majority of the Internet traffic load (file transfer, web browsing, e-mailing, etc.).
- BE—Best Effort. These connections are always admitted, without however any QoS guarantees at all. BE connections share any resources left spare from the superior classes and they gain access to the common channel solely under contention procedures.

Table 1 Available modulation-coding schemes and corresponding SNR thresholds

AMC Index (e)	Modulation & coding	Bits per Symbol ($B(e)$)	Required SNR ($SN_{th,e}$ in dB)
0	–	–	–
1	BPSK & 1/2	0.5	6.4
2	QPSK & 1/2	1	9.4
3	QPSK & 3/4	1.5	11.2
4	16QAM & 1/2	2	16.4
5	16QAM & 3/4	3	18.2
6	64QAM & 2/3	4	22.7
7	64QAM & 3/4	4.5	24.4

In the framework of this priority scheme, new upper-class connections, i.e. UGS, rtPS, ErtPS, are admitted under the loose condition that the aggregate bandwidth requirements (including both the connections already under service and the candidate one) do not exceed $\mathrm{SR}_{\max}$. No additional precautions are taken, since, any potential deterioration of the flows' AMC level should not affect their own QoS, but the corresponding augmentation of bandwidth demand shall be covered in expense of the nrtPS and BE classes. As a result, the efficiency of the CAC functionality is mainly defined by its ability to optimally manipulate the nrtPS traffic. Specifically, CAC bears the burden of restricting the number of accepted connections to make sure that no nrtPS flow falls below its MRTR for more than $q\%$ of the flow's duration, where ideally $q \to 0$. In this process, CAC must take into account (a) the

fluctuations of nrtPS bandwidth demands due to the changes in the terminals' channel conditions as well as (b) any excess of bandwidth demands originating from the higher priority classes, since the UGS, rtPS and ErtPS requirements are bound to be guaranteed at any cost. At the same time, CAC's policy cannot be excessively strict, so as to facilitate the servicing of the largest possible number of nrtPS connections, in order to maximize the network's revenue.

4 CAC: detailed description

As it becomes apparent from Table 1, due to the AMC functionality, the spectrum that needs to be allocated to a connection of specific datarate can range up to a factor of 900%. Therefore, considering also the terminals' mobility, the task of calculating the exact bandwidth claims for the upcoming states of the network is far from being regarded as trivial. In this framework, considering the M-WiMAX distinctive features described in Sect. 3 and extending the research trends presented in Sect. 2, the herein introduced paper aims at augmenting the performance of the CAC's routine by establishing two novel algorithms for the precise computation of the data-flows' resource requirements. In detail, the analysis is performed on both aggregate and per connection basis, using statistical and sampling tools respectively. Eventually, for maximum efficiency, the two parallel BRE (Bandwidth Requirements Estimation) methods are integrated into one hybrid CAC solution.

4.1 Method A: system's average behaviour

The first BRE method calculates the downlink bandwidth requirements per BS as an aggregate, without dealing with the behaviour of every connection individually. The study utilizes the statistical analysis of the overall MSs' spatial distribution and channel shadowing effect, in order to estimate the probability under which each MCS is expected to be engaged by a random connection.

In particular, let $\mathbf{H} \equiv \{1, \ldots, H\}$ be the set of data-flows serviced by the system at time instant t. Then, the mean value of their aggregate bandwidth requirements (averaged throughout their duration) is anticipated to be equal to

$$\mathrm{SR}_{\mathbf{H},\mathrm{av}} = \left\{ P[\mathrm{MC}=0]\cdot 0 + \sum_{e=1}^{E} \frac{P[\mathrm{MC}=e]}{B(e)} \right\} \cdot \sum_{h=1}^{H} \mathrm{DR}_{\mathrm{h}} \quad (3)$$

where $P[\mathrm{MC}=e]$ denotes the probability that a random transmission takes place under the eth MCS. Furthermore, according to the AMC functionality, the probability of engaging a certain MCS ($P[\mathrm{MC}=e]$) coincides with the probability that the SNR value at the receiver end (SN) falls within the corresponding SNR thresholds presented in Table 1. Thus,

$$P[\mathrm{MC}=e] = \begin{cases} P[\mathrm{SN} < \mathrm{SN}_{\mathrm{th},e+1}], & e=0 \\ P[\mathrm{SN}_{\mathrm{th},e} \le \mathrm{SN} < \mathrm{SN}_{\mathrm{th},e+1}, & 1 \le e \le 6 \\ P[\mathrm{SN} \ge \mathrm{SN}_{\mathrm{th},e}], & e=7 \end{cases} \quad (4)$$

Moreover, SN (in dB) is computed as a function of the transmitting power (S_{Tx}), the cumulative propagation losses (A_{ag}) and the channel noise (N_0) that is considered to be Additive White Gaussian Noise (AWGN).

$$\mathrm{SN} = S_{\mathrm{Tx}} - A_{\mathrm{ag}} - N_0 \quad (5)$$

For simplicity in the mathematical analysis, let us denote the constant factor $S_{\mathrm{Tx}} - \mathrm{SN}_{\mathrm{th,e}} - N_0$ with L_e, i.e.

$$L_e = S_{\mathrm{Tx}} - \mathrm{SN}_{\mathrm{th,e}} - N_0 \quad (6)$$

Then, (4) becomes

$$P[\mathrm{MC}=e] = \begin{cases} P[A_{\mathrm{ag}} > L_{e+1}], & e=0 \\ P[L_{e+1} < A_{\mathrm{ag}} \le L_e], & 1 \le e \le 6 \\ P\left[A_{\mathrm{ag}} \le L_e\right], & e=7 \end{cases} \quad (7)$$

Nevertheless, the term A_{ag} can be further simplified, taking into account the fact that the total losses through the propagation path are built up as the sum of three independent terms (Fig. 2) [18]:

1. Distance losses (A_{d}). It principally represents the free-path attenuation and it is directly related to the BS-MS distance (d). Obviously, any alterations of the A_{d} are determined by the terminal's mobility pattern. In the past, several research activities dealt with the issue of forming an as accurate and as generic as possible model for calculating A_{d} as a function of d and the carrier frequency [18, 20]. Based on these models and on the fact that, worldwide, the licensed spectrum allocations for M-WiMAX communication have been made at the 2.3, 2.5, 3.3 and 3.5 GHz frequency bands [3], A_{d} is computed by

$$A_{\mathrm{d}} = f(d) = 128 + 37.6 \cdot \log_{10} d \quad (8)$$

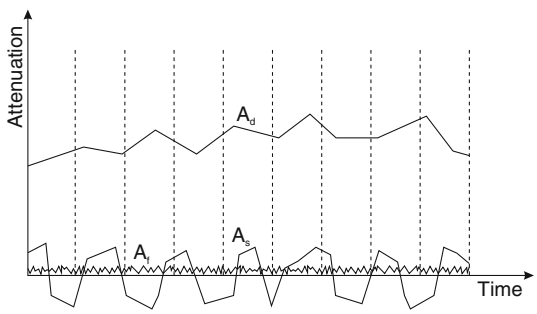


Fig. 2 Aggregate attenuation in wireless propagation environment

where d is measured in km [21]. As it can be easily noticed, A_d is the prevailing attenuation factor in terms of magnitude (circa 128 dBs for MS-BS distance of 1 km). Moreover, it is also very slowly time-varying, since, even in the worst case scenario, a vehicle that travels with a speed of 50 km/h (highest speed allowed within an inhabited area) covers only 7 cm at each timeframe duration.

2. Large-scale shadowing (A_s). It appears due to terrain, building or foliage obstructions that result in absence of Line-of-Sight (LoS) between the transmitter and the receiver. It has been found to follow normal distribution (in dB metrics) with mean value $\overline{A_s} = 0$ dB and standard deviation (σ_{A_s}) that varies from 6 to 12 dB, depending upon the specific propagation environment (urban, suburban, rural etc.). Moreover, as far as the evolution of shadowing against time is concerned, it must be noted that there is significant correlation between successive A_s samples, i.e. the impact of shadowing phenomena spans several consecutive M-WiMAX timeframes without experiencing rather intense A_s fluctuations [18, 22, 23].
3. Small-scale fading (A_f). It is caused by multipath and moving scatterers and it is usually approximated by Rayleigh, Rice, Nakagami-m and Weibull distributions. The SNR variations related to A_f are abrupt and of rather short duration. Furthermore, according to the existing studies, fast-fading causes alterations of the received signal within half a wavelength. Therefore, in the general case, the impact of fast-fading can be safely regarded to be completely uncorrelated between two successive M-WiMAX timeframes (0.005 s) [18, 22].

Taking into account the aforementioned features of each one of these three attenuation constituents, it becomes evident that the evolution of a connection's MCS against time is primarily dictated by free-path losses and shadowing. Specifically, due to the fact that A_f level is supposed to be fully uncorrelated between sequential timeframes, the MCS cannot be chosen solely upon the SNR of a single timeframe, since every SNR measurement is expected to become obsolete within a timeframe's duration. Therefore, in order to filter out the impact of such sporadic, short-lasting phenomena, e.g. small-scale fading, that would cause unreasonable oscillation of the MCS assignment, the output of the AMC mechanism is based on multiple SNR measurements. As a matter of fact, such an averaging method is very common in measurement-based resource management schemes in wireless networks, so as to mitigate the effect of any misleading samples [24]. Moreover, as an additional precaution for ensuring the system's stability, the MCS of each transmission is kept constant throughout every timeframe, i.e. the MCS cannot change more frequently than one timeframe's duration [2]. Thus, there is no point in studying the exact A_f alterations for defining the probability of engaging a specific MCS; only the mean value of the small-scale fading would need to be taken into consideration, since this could indeed shift the average level of the measured SNR. However, in the most common approach of a Rayleigh fading channel [G05] [SA04] [NH05], the probability density function pdf(P_r) of the received power is given by the equation

$$\mathrm{pdf}(P_r) = \frac{1}{\overline{P_r}} \cdot e^{\frac{-P_r}{\overline{P_r}}} \tag{9}$$

where $\overline{P_r}$ is the average value of the received power due to only free path loss and shadowing. As it becomes apparent, the received power after the fast-fading impact is an exponential distribution with mean value $\overline{P_r}$, i.e. the mean alteration factor due to Rayleigh fading is equal to 1 (0 dB) [6, 22, 23]. In addition, it should not be overlooked the fact that, at significant level, the effect of A_f has already been incorporated in the process of determining the AMC's SNR thresholds so as to ensure that it is drastically mitigated by the diversity, equalizing, coding and interleaving schemes. On the contrary, the A_d and A_s variations, which are triggered by the MS's movement, introduce severe and long-lasting SNR fluctuations (high correlation between successive A_d and A_s samples) and hence these are the ones that play the predominant role in determining the transmission's MCS. As a result, it can be safely regarded that

$$A_{ag} = A_d + A_s \tag{10}$$

At this point it must be underlined that the present BRE method aims at formulating a reference model for calculating the expected distribution of a given set of connections among the available MCSs, so as to determine their mean resource requirements. In this framework, although small-scale fading may temporarily cause significant degradation of a data-flow's quality (erroneous packet reception), it cannot affect the average ratio of connections engaging a specific MCS over a substantial amount of time; thus the impact of A_f does not need to be examined on per timeframe basis. Hence, using Eq. 10 is a common practice for estimating the mean capacity of an access network [8, 18]. In any case, as it will be presented below, BRE method A introduces also a calibration routine for the exact purpose of incorporating all the possible variations of the actual operating system in comparison with the reference model.

Based on Eq. 10 and considering that A_s follows Gaussian distribution, as it was previously described, the distribution of A_{ag} in its general format is depicted in Fig. 3. For a given value of $A_d, A_d = A_{d_x}$,

$$P\left[L_{e+1} < A_{\text{ag}} \le L_e\right] = P[L_{e+1} - A_{d_x} < A_s \le L_e - A_{d_x}] = \int_{L_{e+1}-A_{d_x}}^{L_e - A_{d_x}} \text{pdf}(A_s)dA_s \quad (11)$$

where pdf(A_s) is the probability density function of the variable A_s:

$$\text{pdf}(A_s) = \frac{1}{\sigma_{A_s}\sqrt{2\pi}} \cdot \exp\left\{\frac{-(A_s - \overline{A_s})^2}{2\sigma_{A_s}^2}\right\} \quad (12)$$

and the general notation $P[\ldots]$ shall hereafter denote the probability that the condition within the brackets is true.

However, A_d is not a static parameter but it is also a random variable, which takes values within the space $[A_{d_m}, A_{d_M}]$ ($A_d \in [A_{d_m}, A_{d_M}]$). A_{d_m} and A_{d_M} denote the distance-attenuation experienced respectively at the locations of minimum (d_m) and maximum (d_M) BS-MS distance, depending on the topology of the M-WiMAX access network (radius of a M-WiMAX cell). Therefore,

$$\left[L_{e+1} < A_{\text{ag}} \le L_e\right] = \int_{A_{d_m}}^{A_{d_M}} P[A_d = A_{d_x}] \left(\int_{L_{e+1}-A_{d_x}}^{L_e - A_{d_x}} \text{pdf}(A_s)dA_s \right) dA_{d_x} \quad (13)$$

Judging from Eq. 13, in order to calculate the probability of engaging the eth MCS ($P[\text{MC} = e]$), it is necessary to define the distribution of A_d. For this purpose, we adopt the common assumption that, on average throughout the timeline, the whole set of MSs generating the data connections under service are uniformly distributed within the boundaries of the M-WiMAX cell. Moreover, since the integral in Eq. 11 cannot be analytically resolved as a function of A_{d_x} (the probability distribution function of a Gaussian variable cannot be analytically computed), the so far continuous space of the A_d values ($A_d \in [A_{d_m}, A_{d_M}]$) is quantised with step $a \longrightarrow 0$ into a very large number (b) of discrete, sequential values that are marked as A_{d_x}. Hence, hereafter,

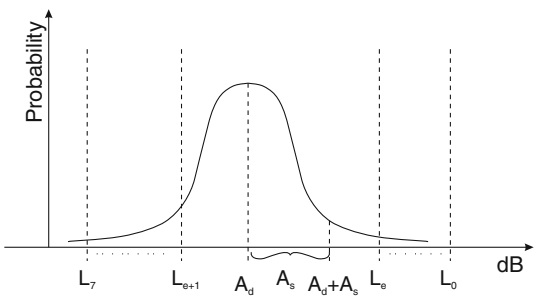


Fig. 3 Distribution of A_{ag}

$$A_{d_x} \in \{A_{d_m}, A_{d_m} + a, \ldots, A_{d_m} + (b-2)a, A_{d_M}\} \quad (14)$$

with $x = \{1, \ldots, b\}$. Similarly,

$$P\left[L_{e+1} < A_{\text{ag}} \le L_e\right] = \sum_{x=1}^{b} \left\{ P[A_d = A_{d_x}] \left(\int_{L_{e+1}-A_{d_x}}^{L_e - A_{d_x}} pdf(A_s)dA_s \right) \right\}$$

Furthermore, since A_d is a variable of discrete type, by definition we have that

$$P[A_d = A_{d_x}] = P[A_d \le A_{d_x}] - P\left[A_d \le A_{d_x}^-\right] \quad (15)$$

and due to the fact that the quantising step $a = (A_{d_x} - A_{d_{x-1}})$ is extremely small,

$$P[A_d = A_{d_x}] = P[A_d \le A_{d_x}] - P[A_d \le A_{d_{x-1}}] \quad (16)$$

In this respect, considering a circular M-WiMAX cell of radius r_c ($r_c = d_M$), the BS coverage area can be split into b concentric circles of radius $r_x = f^{-1}(A_{d_x}) \in [d_m, d_M]$ (Fig. 4), where $A_d = f(d)$ is described in Eq. 8 as the formula according to which A_d is computed as a function of the BS-MS distance d. As a result, the probability that $A_d \le A_{d_x}$ coincides with the probability that $d \le r_x$, i.e.

$$P[A_d \le A_{d_x}] = P[d \le r_x] \quad (17)$$

Moreover, taking into account that the terminals' spatial distribution is expected to be uniform, the probability that an MS is located within a certain ring is equal to the ratio of this ring's area to the aggregate coverage area (from d_m to d_M), i.e.

$$P[d \le r_x] = \frac{r_x^2 - d_m^2}{r_c^2 - d_m^2} \quad (18)$$

Thus, combining Eqs. 16, 17 and 18 for $a \longrightarrow 0 \Leftrightarrow A_{d_{x-1}} \longrightarrow A_{d_x}^-$, it turns out that

$$P[A_d = A_{d_x}] = \frac{(f^{-1}(A_{d_x}))^2 - (f^{-1}(A_{d_{x-1}}))^2}{r_c^2 - d_m^2} \quad (19)$$

Summarizing all the previous equations, the requested probabilities $P[\text{MC} = e], \forall e \in \{0, 1, \ldots, 7\}$ can be computed from

$$P[\text{MC} = e] = \begin{cases} \sum_{x=1}^{b} \left\{ P[A_d = A_{d_x}] \left(\int_{L_{e+1}-A_{d_x}}^{+\infty} pdf(A_s)dA_s \right) \right\}, & e = 0 \\ \sum_{x=1}^{b} \left\{ P[A_d = A_{d_x}] \left(\int_{L_{e+1}-A_{d_x}}^{L_e - A_{d_x}} pdf(A_s)dA_s \right) \right\}, & 1 \le e \le 6 \\ \sum_{x=1}^{b} \left\{ P[A_d = A_{d_x}] \left(\int_{-\infty}^{L_e - A_{d_x}} pdf(A_s)dA_s \right) \right\}, & e = 7 \end{cases} \quad (20)$$

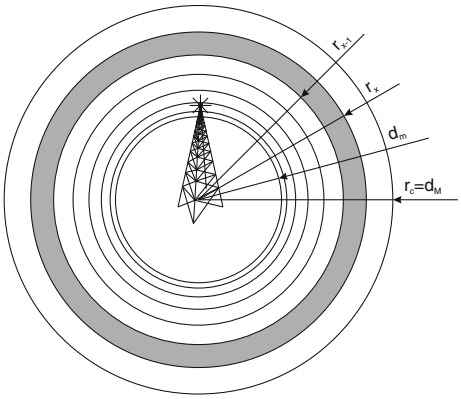


Fig. 4 Analysis of BS coverage area into concentric circles

where pdf(A_s) and $P[A_d = A_{d_x}]$ are defined in Eqs. 12 and 19, respectively. Eventually, by incorporating Eq. 20 into Eq. 3, we calculate the average symbol-rate requirements of a given set of data-flows (H).

Nevertheless, as it becomes evident, the above results are reached via the probabilistic analysis of stochastic phenomena such as distance-attenuation, shadowing, noise etc. under certain assumptions. In more detail, the study makes virtue of the knowledge of the available probability distribution parameters, i.e. uniform spatial distribution for the MSs, Gaussian distribution for the shadowing effect and AWGN for the channel noise. In this framework, in order such a statistical approach to be applicable to the whole variety of propagation environments, a suitable calibration process is introduced. In particular, let $\overline{\mathrm{SN}_{\mathrm{st}}}$ be the mean value of SNR that is statistically expected based on the proposed scheme. Using (5) and (10), $\overline{\mathrm{SN}_{\mathrm{st}}}$ is equal to:

$$\overline{\mathrm{SN}_{\mathrm{st}}} = S_{T_x} - \overline{A_d} - \overline{A_s} - N_0 = S_{T_x} - \overline{A_d} - N_0 \tag{21}$$

where

$$\overline{A_d} = \sum_{x=1}^{b} \{P[A_d = A_{d_x}] A_{d_x}\} \tag{22}$$

Furthermore, let $\overline{\mathrm{SN}_{\mathrm{mr}}}$ be the mean value of SNR that is actually being measured by the specific BS within a loose time-window equal to the connections' average duration. Then, the calibration factor V is defined as the difference between $\overline{\mathrm{SN}_{\mathrm{st}}}$ and $\overline{\mathrm{SN}_{\mathrm{st}}}$:

$$V = \overline{\mathrm{SN}_{\mathrm{st}}} - \overline{SN_{\mathrm{mr}}} \tag{23}$$

Finally, Eq. 20 is computed for

$$L'_e = L_e - V \tag{24}$$

As it has already been pointed out, the present BRE method aims at pinpointing the average spectrum requirements of the whole system (or equivalently the average capacity of the access network), without following the evolution of each data-connection separately. Hence, the notion behind the introduction of the calibration routine is that V should be representative of any (semi)permanent variation between the aggregate system under study ($\overline{\mathrm{SN}_{\mathrm{mr}}}$) and the prototype one ($\overline{\mathrm{SN}_{\mathrm{st}}}$). This way, the utilization of V provides the ability to seamlessly incorporate all the unpredictable factors that may cause this (semi)permanent shift in the foreseen SNR distribution, such as non-uniform spatial distribution of the terminals due to certain terrain peculiarities/obstacles, interference from neighbouring WiMAX cells (if any), the mean value of the fast-fading when it becomes of notable level etc.

Concluding, regarding any implementation complications, it must be mentioned that the proposed BRE scheme causes no latency overhead, since, not only the processing complexity of the averaging algorithm is negligible, but all the necessary calculations are executed off-line as well. Moreover, apart from the number of concurrent connections and their datarate requirements, which are anyway available at the NCC database, no extra information about the system state is requested. Hence, no additional signalling is exchanged.

Additionally, as far as the anticipated accuracy of the method is concerned, it is expected to be maximized, the higher the number (H) of serviced connections is and the longer these connections last, because, in this case, the observation window (number of connections multiplied by their duration) is large enough in order the behaviour of the actual system to converge with the statistically expected one (mean value, standard deviation etc.). In this manner, Method A is mainly introduced with the aim of providing a relatively long-term estimation of the cumulative MSs' bandwidth demands. Such info is of high importance, as, according to Sect. 3.2, the efficiency in supporting the decisive nrtPS class is evaluated on the basis of the whole connection's duration, due to the flexible delay limitations of these applications, i.e. sporadic deteriorations cannot justify the overallocation of resources.

4.2 Method B: prospective behaviour per connection

In contrast to the cumulative approach of Method A, the second algorithm focuses on the particular behaviour of each data-flow separately. Exploiting the available SNR measurements of the MSs under service, the herein presented BRE mechanism develops a sophisticated scheme for the processing of the SNR info, in order to accurately formulate the bandwidth allocation criteria. The acquired SNR samples are filtered through a heuristic routine so as only the ones that are indicative of the channel's condition to be taken into account [25].

Being more specific, the prediction of the exact amount of resources to be requested by each connection is based on the following principles:

– In general, a given MS does not wanders randomly, but it travels towards a specific destination in conformity with the personalized intentions of its bearer. In detail, except for very seldom cases, an MS can not be alleged to follow a straight line throughout the duration of a data-stream, since it is almost always mandatory for the mobile user to take turns along the available path, i.e. roads for vehicular users and walking paths for pedestrians. Nevertheless, due to the fact that any trajectory variations are bounded by the user's will to approach the target location, it can be safely taken for granted that the distance of each MS from the BS either augments or decreases in an averagely steady manner. Thus, any temporal fluctuations of the MS's SNR should not be taken into account in the prediction of the SS's bandwidth requirements. In the illustrative example of Fig. 5, the terminal starts the data transmission at point A and finishes at point B, following *route*2 (realistic) instead of *route*1 (ideal). Based on the analysis above, any SNR measurements taken at points C or D should be discarded, since they are not indicative of the MS's average state.
– As it has already been thoroughly described in the previous subsection (Method A) and summarised in Eq. 10, the aggregate degradation (A_{ag}) of the transmitted signal is mainly caused by two factors: a) the attenuation due to distance (A_d) and b) the fast-fading phenomenon (A_f) because of temporal absence of LOS (Line Of Sight). As a result, any short-term SNR changes can be regarded to be caused by the shadowing component and thus they should not be taken into account as they are only transitory and they can be misleading about the future channel conditions.

In order to quantify the above statements, we propose an adequate averaging scheme, capable of discarding any attenuation peaks as well as quickly adopting to the gradual change of the MS-BS distance, so as to make safe predictions about the future demands of the admitted MSs. In detail, at the time instant t of a new connection arrival, the proposed CAC scheme initially performs the statistical analysis of the available SNR samples that are taken at steady intervals of T_{smp} seconds:

1. $\forall\ h \in \mathbf{H}$, for $\mathbf{H} \equiv \{1, \ldots, H\}$ being the set of all the connections being under service at time t, the whole duration of h (beginning from h's start-up) is divided into time windows of the same width (T_{wnd}). Each time window is defined by its index number $1_h, \ldots, k_h, \ldots, K_h$ and t_{k_h} is the end of the k_h time window, i.e. $t_{k_h} - t_{(k-1)_h} = T_{wnd}$ and $t_{(K-1)_h} < t \leq t_{K_h}$ (Fig. 6)
2. Let $N_{smp} = T_{wnd}/T_{smp}$ be the number of samples taken within T_{wnd}. Then, $\forall\ h \in H$, each SNR sample within a given time window $k_h \in \{1_h, \ldots, K_h\}$ is denoted as $SN_{h,k,j}$, where $j \in \{1, \ldots, J_h\}$ and J_h
$$\begin{cases} \leq N_{smp}, & k_h = K_h \\ = N_{smp}, & k_h < K_h \end{cases}$$
3. We introduce Q as the threshold for discarding any excessively high and low SNR samples. The SNR reference level for the SNR samples to be compared against is defined equal to the SNR level of the current time window ($SN_{h,k,av}$). According to this clause, an SNR sample of the kth time window of the hth connection is taken into account if and only if it satisfies both Eqs. 25 and 26:
$$SN_{h,k,j} < SN_{h,k,av} \cdot (100 + Q)\% \quad (25)$$
$$SN_{h,k,j} > SN_{h,k,av} \cdot (100 - Q)\% \quad (26)$$
where $SNR_{h,k,av}$ denotes the mean SNR of the kth sampling window of flow h:
$$\Sigma = SN_{h,k-1,av} + \sum_{j=1}^{J_h} SN_{h,k,j} \quad (27)$$
$$SN_{h,k,av} = \frac{1}{J_h + 1} \cdot \Sigma \quad (28)$$
and $SN_{h,k,j}$ satisfies Eqs. 25 ∩ 26

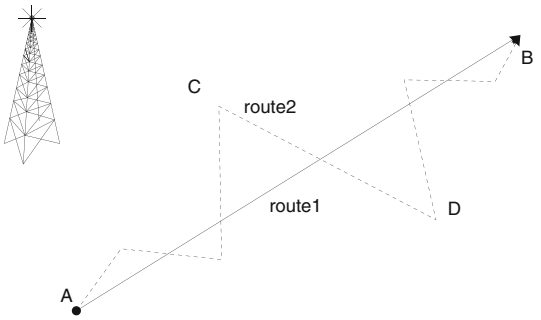


Fig. 5 General mobility pattern in a cellular architecture

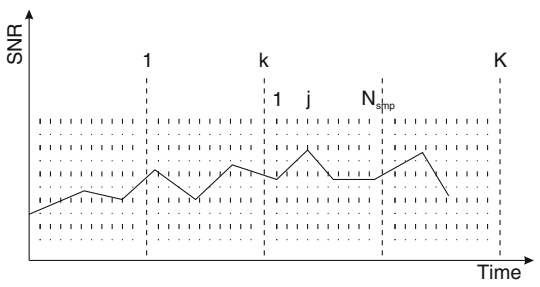


Fig. 6 SNR sampling procedure

At the next step, on the basis of the aforementioned procedure, the CAC algorithm estimates the prospective MCS of all connections ($MC_{h,t,pr}$). In particular, $\forall\ h \in \mathbf{H}$, let $MC_{h,t} = e_h$, i.e. the MCS of the hth flow at time t is equal to $e_h \in \mathbf{E} \equiv \{1, \ldots, 7\}$. The prospective MCSs are calculated as follows:

– $MC_{h,t,pr} = e_h - 1$, if and only if

$$SN_{h,K,av} < SN_{h,K-1,av} \text{ and } SN_{h,K,av,} < SN_{th,e_h} \cdot (100 + X)\%$$

– $MC_{h,t,pr} = e_h + 1$, if and only if

$$SN_{h,K,av} > SN_{h,K-1,av} \text{ and } SN_{h,K,av} > SN_{th,(e_h+1)} \cdot (100 - X)\%$$

– Elsewhere, $MC_{h,t,pr} = e_h$

The above routine for predicting each connection's upcoming MCS is graphically presented in Fig. 7. As it becomes obvious, for a terminal with given $SN_{K,av}$, the probability of upgrading to a higher MCS is slightly greater than the probability of falling to a lower MCS, since the limits $SN_{th,e} \cdot (100 \pm X)\%$ are getting broader for increased values of the SNR threshold ($SN_{th,e}$). This small tendency towards upper MCSs introduces a desirable level of optimism to the CAC policy, which is necessary for further improving resource exploitation.

Utilizing the output of the previous steps $MC_{h,t,pr}$ and taking into account Eq. 2, the predicted aggregate SR requirements of all the currently serviced connections are:

$$SR_{\mathbf{H},pr} = \sum_{h=1}^{H} SR_h = \sum_{h=1}^{H} DR_h \cdot B(MC_{h,t,pr}) \tag{29}$$

As far as the formulation of the proposed sampling-prediction algorithm is concerned, it can be described in brief that the prediction of the upcoming MCS is based on the comparison between the non-weighted average of the SNR samples of each connection's two last time-windows. This particular choice has been made on the basis that, due to the fact that the correlation among the samples of the A_f component is practically zero, such a configuration allows the impact of the abrupt and misleading fast-fading fluctuations to be radically mitigated. Being more specific, within two successive sampling-windows the contribution of the A_f factor to the non-weighted average of the SNR samples is constant, since the mean value of the A_f samples rapidly converges to the mean value of the expected fast-fading distribution (Rayleigh, Rice, Nakagami-m etc.). Hence, the comparison of the non-weighted average SNR of the two consecutive time-windows isolates the trend of the underlying, slow-varying components (A_d and A_s). Additionally, the exclusion of any temporal excessive variations of the SNR level is further supported by the filtering routine that is implemented in Eqs. 25 and 26.

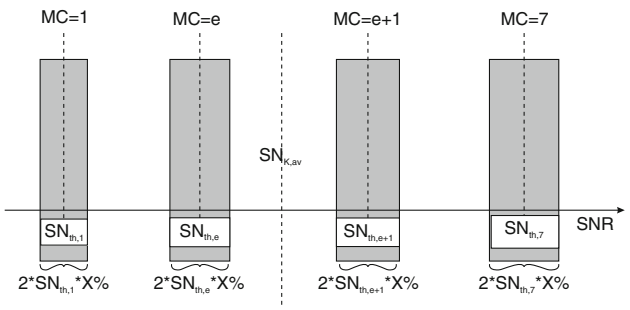


Fig. 7 Routine for the prediction of the upcoming state

Finally, it must be underlined the fact that, alike Method A, the present BRE mechanism imposes no additional delay or signalling overhead, since the SNR information is already available from the AMC routine. Moreover, opposite from the averaging approach of Sect. 4.1, Method B makes no pre-assumptions about either the terminals' spatial distribution and mobility characteristics (e.g. velocity) or the propagation environment. On the contrary, this algorithm aims at performing the solid estimation of each MS's prospective transmission quality, through effectively interpreting the MS's SNR measurements. As a result, such a strategy allows the short-term mapping of every data-flow's future bandwidth requirements. This way, CAC is rapidly adapted to the altering network's status and thus the data-streams' QoS standards are guaranteed on per timeframe basis.

4.3 Hybrid CAC

In Sects. 4.1 and 4.2 we introduced two novel and independent with each other algorithms for accurately predicting the upcoming resource requirements (symbols/sec) of a given set of active connections, taking into account the dynamic alterations of the M-WiMAX physical layer configuration. Furthermore, on the basis of this analysis, we propose a hybrid CAC scheme that amalgamates the above BRE outcomes in order to formulate a unified admittance/rejection decision.

Being more specific, let $\mathbf{H}$ be the set of active connections at the time instant t when a new data-flow applies for service. Provided the datarate limitations (DR_{cc}) and the contemporary AMC state ($MC_{cc} \in \mathbf{E}$) of the candidate connection (cc), the admittance decision criterion can be threefold:

– CAC based on BRE method A (CAC_A):

$$SR_{\mathbf{H},av} + \frac{DR_{cc}}{B(MC_{cc})} \leq SR_{max} \tag{30}$$

– CAC based on BRE method B (CAC_B):

$$\mathrm{SR}_{\mathbf{H},\mathrm{pr}} + \frac{\mathrm{DR}_{\mathrm{cc}}}{B(\mathrm{MC}_{\mathrm{cc}})} \leq \mathrm{SR}_{\max} \tag{31}$$

– Proposed Hybrid CAC ($\mathrm{CAC}_{\mathrm{HY}}$):

$$\min\left\{SR_{\mathbf{H},\mathrm{av}}, \mathrm{SR}_{\mathbf{H},\mathrm{pr}}\right\} + \frac{\mathrm{DR}_{\mathrm{cc}}}{B(\mathrm{MC}_{\mathrm{cc}})} \leq \mathrm{SR}_{\max} \tag{32}$$

where $\mathrm{SR}_{\mathrm{H,av}}$ and $\mathrm{SR}_{\mathrm{H,pr}}$ are computed from (3) and (29), respectively.

Judging from Eq. 32, the hybrid CAC implements both BRE Method A and BRE Method B in parallel and subsequently, according to their outputs, it considers the minimum traffic load that is anticipated to be generated by the set of the already accepted connections (**H**). The incorporation of two supplementary perspectives for the resolution of the capacity estimation problem guarantees the optimization of the CAC's performance, since the integrated CAC mechanism embodies all the favourable features of its dual functionality. Moreover, the rational behind choosing the most optimistic, per occasion, scenario ($\min\left\{\mathrm{SR}_{\mathbf{H},\mathrm{av}}, \mathrm{SR}_{\mathbf{H},\mathrm{pr}}\right\}$) is justified by the ultimate goal of maximizing the utilization of the scarce spectrum resources through narrowing down any over-allocation phenomena.

5 Simulation and performance evaluation

5.1 Optimization criteria

Based on the analysis provided in Sects. 3 and 4, two metrics are specified so as to quantitatively measure the effectiveness of the CAC mechanism:

– Blocking Probability (BP). The ratio of the number of blocked connections against the aggregate number of connections' arrivals to the system. It is indicative of the CAC's capability to maximize network's revenue by fully exploiting the system's bandwidth.
– Mean Satisfaction Factor (MSF). According to Sect. 3.2, the DR of each rtPS or nrtPS connection should not fall below MRTR and it cannot exceed MSTR. However, in cases of resource shortage because of unexpectedly excessive symbol-rate requests (over-optimistic CAC policy), a data-flow's DR shall be analogously degraded at an extent that is dependent on its QoS class priority privileges. In this respect, SF_{g} is defined as the average of the $\mathrm{DR}_{\mathrm{g}}/\mathrm{MRTR}_{\mathrm{g}}$ fraction for the whole duration of the gth connection, where $g \in \mathbf{G}$ and $\mathbf{G} \equiv \{1, \ldots, G\}$ is the set of all the connections that have been serviced from the beginning of the system's study. Then, MSF is equal to the mean value of all the recorded SF values and it is representative of the CAC's ability to fulfil the network's QoS expectations.

$$\mathrm{MSF} = \frac{1}{G} \cdot \sum_{g=1}^{G} \mathrm{SF}_{\mathrm{g}} \tag{33}$$

As it becomes evident, the optimization problem is summarized in minimizing BP, while keeping $\mathrm{MSF} \geq 1$ ($\mathrm{BP} \longrightarrow 0 \cap \mathrm{MSF} \longrightarrow 1^{+}$).

5.2 Reference schemes

In order to comparatively evaluate the ability of the proposed hybrid CAC algorithm ($\mathrm{CAC}_{\mathrm{HY}}$) to meet the revenue (BP) and QoS (MSF) criteria, $\mathrm{CAC}_{\mathrm{HY}}$ is tested against two existing and generally implemented CAC algorithms. In particular, similarly to $\mathrm{CAC}_{\mathrm{HY}}$, both CAC algorithms of reference take into account the AMC feedback. However, they differ with each other as well as with the $\mathrm{CAC}_{\mathrm{HY}}$ regarding the manipulation of the physical layer info. Specifically:

1. The first of the reference CAC algorithms ($\mathrm{CAC}_{\mathrm{R1}}$) performs no statistical process of the SNR samples. The estimation of the necessary bandwidth for each connection is based solely on the instantaneous SNR measurement.
2. The second reference CAC scheme ($\mathrm{CAC}_{\mathrm{R2}}$) buffers the last $N_{\mathrm{R2,smp}}$ SNR samples so as to calculate the prospective MCS based on the average value of these samples.

5.3 Simulation model and parameters

The performance of the CAC procedure is assessed through a simulation software that has been developed by the authors in C++ and it is capable of realistically recreating the network environment as well as implementing the CAC algorithms. The reference architecture considers the downlink channel of a single-cell M-WiMAX system, with radius $r_c = d_M = 1.5$ km and minimum distance $d_{\mathrm{m}} = 20$ m. Every terminal moves randomly within the cell boundaries with a constant speed s, where s is uniformly distributed between 4 and 50 km/h (from urban pedestrians to suburban vehicular users). The rest of the physical layer parameters of the simulated IEEE 802.16e system, as these are defined in [2, 3], are summarized in Table 2.

Two QoS classes, rtPS and nrtPS, are considered as well as two different MRTR levels, MR1 = 128 kbps and MR2 = 256 kbps. The ratio of the rtPS to the nrtPS connections is 1/3, while the set of simulated data-flows is evenly split between the two MRTR levels (P[M RT $\mathrm{R_g}$ = M R1] = P[M RT $\mathrm{R_g}$ = M R2] = 50%, $\forall\ g \in$ G). Moreover, the connections' arrival process follows a Poisson behaviour, i.e. both the connections' interarrival time (I) and

duration (D) are exponentially distributed with mean values equal to $\overline{I}$ and $\overline{D}$ respectively. Being more specific, in order to enhance the realism of the simulation scenario, the rtPS data-flows are supposed to have an exponentially distributed duration (D_{rt}) with mean value $\overline{D_{\mathrm{rt}}} = 300s$, while the duration of the non real-time connections $\overline{D_{\mathrm{nrt}}}$ is considered to be limited by the data-size (S_{nrt}) of the file that is assumed to be forwarded, which is also exponentially distributed with mean value $\overline{S_{\mathrm{nrt}}} = 32Mb$.

Finally, in order to be able to simulate disparate traffic conditions, we introduce $\mathrm{DR}_{\mathrm{ag,nrm}}$ as the average rate of the aggregate traffic that is being directed into the network, normalised to the network's maximum capacity. In detail, let $\mathrm{DR}_{\mathrm{max}}$ be the maximum datarate that can be forwarded by the system

$$\mathrm{DR}_{\mathrm{max}} = \frac{C_{\mathrm{dat}} \cdot L_{\mathrm{tot}} \cdot B(7)}{T_{\mathrm{frm}}} = 27.216(\mathrm{Mbps})$$

and let $\overline{\mathrm{DR}_{\mathrm{ag}}}$ be the mean value of the cumulative bitrate requirements at each time instant

$$\begin{aligned}\overline{\mathrm{DR}_{\mathrm{ag}}} &= \frac{\overline{D}}{\overline{I}} \cdot (0.5 \cdot MR1 + 0.5 \cdot MR2) \\ &= \frac{0.25\overline{D_{rt}} + 0.75\overline{D_{\mathrm{nrt}}}}{\overline{I}} \cdot 192(kbps) \\ &= \frac{0.25 \cdot 300 + 0.75 \cdot \frac{32 \cdot 10^3}{192}}{\overline{I}} \cdot 192(\mathrm{kbps}) \\ &= \frac{38.4}{\overline{I}}(\mathrm{Mbps})\end{aligned} \tag{34}$$

Then, $\mathrm{DR}_{\mathrm{ag,nrm}}$ is defined as:

$$\mathrm{DR}_{\mathrm{ag,nrm}} = \frac{\overline{\mathrm{DR}_{\mathrm{ag}}}}{\mathrm{DR}_{\mathrm{max}}} \tag{35}$$

The mean interarrival time $\overline{I}$ (in seconds) is left open, so as to be used as the regulator of the traffic intensity.

Table 2 Physical layer simulation parameters

Parameters of downlink 802.16e	Value
Total BS transmission power ($S_{\mathrm{Tx,ag}}$)	20 W
Total bandwidth (W_{tot})	10 MHz
Total number of OFDMA subcarriers (C_{tot})	1,024
Number of OFDMA subcarriers for data transmission (C_{dat})	720
Number of timeslots per frame (L_{tot})	42
Duration of a time-frame (T_{frm})	0.005 s
Standard deviation of shadowing (σ_{A_s})	8 dB
Gaussian noise power (N_0)	$10^{(-174/10)}$ mW/Hz
DL/UL transmission mode	TDD

5.4 Definition of CAC's optimum configuration

Before proceeding with the comparative evaluation of the hybrid CAC algorithm, the optimal configuration of both the proposed ($\mathrm{CAC}_{\mathrm{HY}}$) and the reference ($\mathrm{CAC}_{\mathrm{R1}}$ and $\mathrm{CAC}_{\mathrm{R2}}$) CAC schemes under the specific scenario must be pinpointed, so as the credibility and fidelity of the results to be guaranteed. Specifically, the computation of $\mathrm{SR}_{\mathrm{H,pr}}$, which, as described in Eq. 32, forms the first of the two pillars of the hybrid CAC decision, introduces three new parameters: the extent of the connections' history that needs to be taken into account (N_{smp}), the threshold for discarding the misleading SNR samples (Q) and the threshold for predicting the upcoming AMC state (X). Similarly, the behaviour of $\mathrm{CAC}_{\mathrm{R2}}$ is expected to be bounded by the width of the averaging window (N_{R2},smp).

In this framework, the performance of the $\mathrm{CAC}_{\mathrm{HY}}$ and $\mathrm{CAC}_{\mathrm{R2}}$ algorithms are simulated under the whole range of the aforementioned variables, for the medium traffic load of $\mathrm{DR}_{\mathrm{ag,nrm}} = 0.44$, and the results are presented in Figs. 8, 9, 10 and 11. As a matter of fact, in order to isolate exclusively the impact of the N_{smp}, Q and X parameters among the rest of the hybrid CAC's factors, instead of testing the $\mathrm{CAC}_{\mathrm{HY}}$, we measure the BP and MSF of the $\mathrm{CAC}_{\mathrm{B}}$ Eq. 31.

In general, the efficiency of sampling-based CAC strategies, such as $\mathrm{CAC}_{\mathrm{B}}$ and $\mathrm{CAC}_{\mathrm{R2}}$, increases the greater the amount of available samples is, since the statistical accuracy is maximized. On the other hand, their complexity is directly proportional to the bulk of the data that is manipulated, i.e. the number of SNR samples that have to be stored and processed. As a result, interpreting Figs. 8 and 9, $\mathrm{CAC}_{\mathrm{B}}$ needs to maintain a history of at least 10 SNR samples ($N_{\mathrm{smp}} \geq 10$) so as its performance to rise up to the desirably enhanced levels. Such an output can be considered as completely natural, since, according to Eqs.25–28, the proposed algorithm requires a substantial amount of info so as to built the SNR level of reference against which all the SNR samples are compared. On the contrary, $\mathrm{CAC}_{\mathrm{R2}}$ reaches a steady state for the minimum number of 3 SNR samples ($N_{\mathrm{R2,smp}} \geq 3$), while for $N_{\mathrm{R2,smp}} = N_{\mathrm{smp}} = 1$ both $\mathrm{CAC}_{\mathrm{HY}}$ and $\mathrm{CAC}_{\mathrm{R2}}$ match the basic $\mathrm{CAC}_{\mathrm{R1}}$ CAC algorithm (instantaneous monitoring). Despite this incrementation in sampling history needs ($N_{\mathrm{smp}} = 10$), $\mathrm{CAC}_{\mathrm{B}}$ is guaranteed to be completely functional and scalable, since the corresponding time window is still kept very low. In detail, the necessary sampling duration is $T_{\mathrm{wnd}} = N_{\mathrm{smp}} \cdot T_{\mathrm{smp}} = 0.5s$, which means that every terminal has to remain within a BS's coverage for at least 0.5 s, i.e. the maximum inter-handover time cannot exceed 0.5 s. However, such a

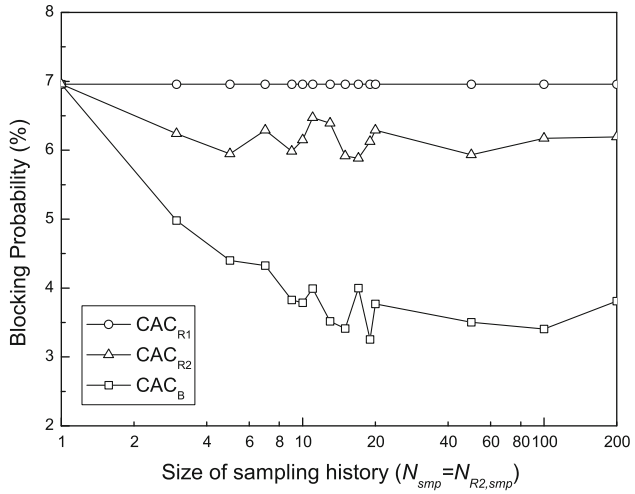


Fig. 8 Blocking Probability for different sizes of sampling window

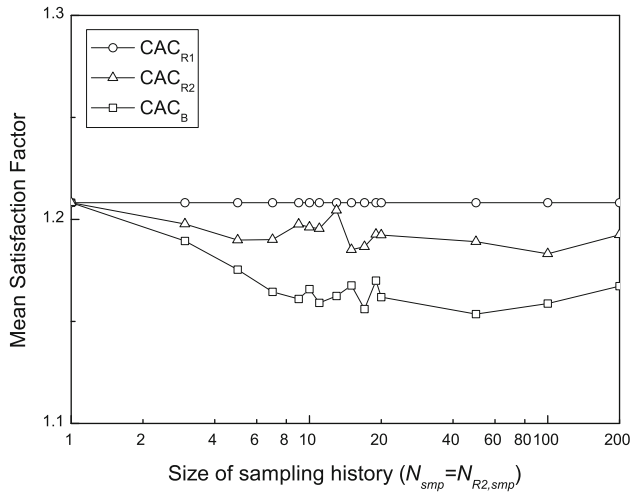


Fig. 9 Mean Satisfaction Factor for different sizes of sampling window

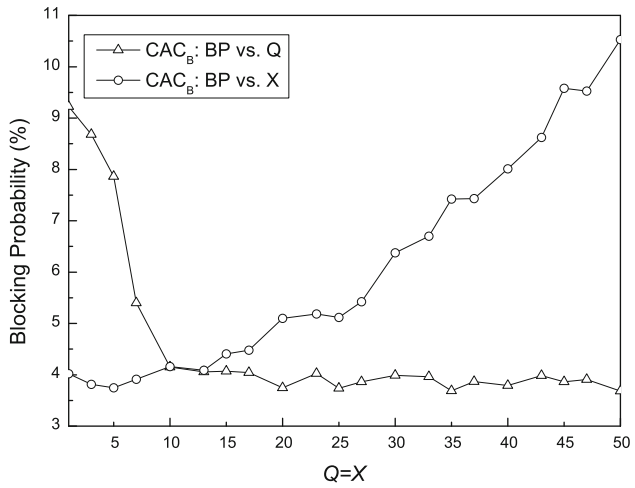


Fig. 10 Blocking Probability for different values of Q and X

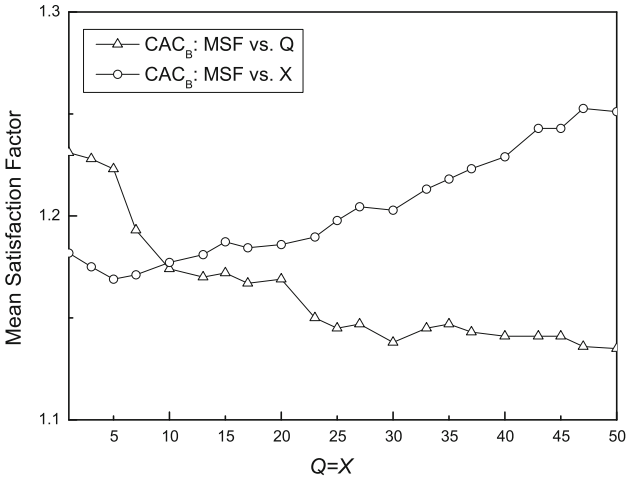


Fig. 11 Mean Satisfaction Factor for different values of Q and X

limitation is always satisfied, even for the case of the smallest cells.

Furthermore, as it has been established in Sect. 4.2, Q determines the range of SNR samples that are taken into account, i.e. the lower is Q the higher is the percentage of SNR samples that are discarded as being non-representative of the connection's general pattern. Hence, it is expected that the algorithm's efficiency would be degraded for the marginal values of Q due to the rather strict ($Q \to 0$) or loose ($Q \to \infty$) filtering process. Accordingly, based on the numerical results of Figs. 10 and 11:

- BP is a genuinely descending function for $0.01 \leq Q \leq 10$, while it seems to remain practically unaltered for $Q \geq 10$. Thus, as far as the BP is concerned, any value of $Q \geq 10$ is acceptable.
- MSF is also a genuinely descending function of Q, while it undergoes an additional significant degradation for $Q > 20$. Hence, regarding MSF behaviour, any value of $Q \leq 20$ is acceptable.

However, among all the values within the favouring space $10 \leq Q \leq 20, BP$ presents the minimum value for $Q = 20$. Therefore, it can be derived the conclusion that the optimal balance between the competing BP and MSF behaviour is achieved for $Q = 20$.

On the other hand, X denotes the maximum acceptable variation from an MCS's SNR threshold so as this MCS to be regarded as the upcoming one. Hence, there would be no point in choosing a large value for X, since in that case there would be an overlap with the SNR thresholds of the neighbouring MCSs (Table 1). On the contrary, very small values of X would practically lead to disabling the prediction module. As a result, judging from Figs. 10 and 11, the value of $X = 5$ is chosen.

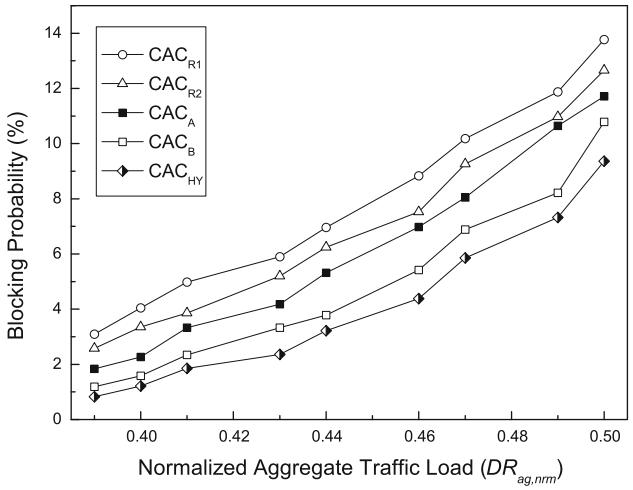


Fig. 12 Blocking Probability for different traffic conditions

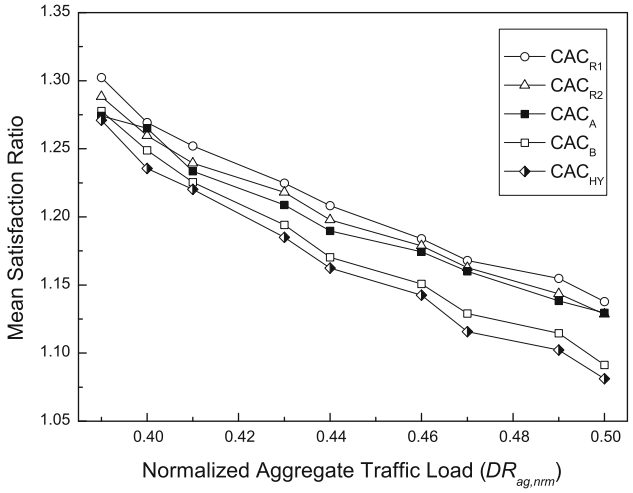


Fig. 13 Mean Satisfaction Factor for different traffic conditions

5.5 Comparative Study

In the next step, for the aforementioned set of optimum values ($Q = 20$, $X = 5$, $N_{\text{smp}} = 10$ and $N_{\text{R2,smp}} = 3$), the efficiency of the hybrid CAC scheme is tested under different magnitudes of traffic load. The simulation output is presented in Figs. 12 and 13, which also illustrate the autonomous performance of CAC_A and CAC_B, so as to be able to quantify the separate contribution of the two individual components of the hybrid CAC towards the overall scheme's effectiveness. For further detail, the comparative improvement achieved by the utilization of CAC_HY in respect to the CAC_R2 and CAC_R1 performance is calculated in Table 3. Although, the theoretical calculations as well as the simulations have been carried out for a much broader range of traffic loads, only the results that correspond to the operational range of $1\% \leq \text{BP} \leq 10\%$ are herein shown.

Table 3 BP improvement achieved by the proposed hybrid CAC algorithm

$\text{DR}_{\text{ag,nrm}}$	BP Improvement of CAC_HY against CAC_R1 (%)	BP Improvement of CAC_HY against CAC_R2 (%)
0.39	73.19	67.82
0.40	70.07	63.91
0.41	62.87	52.02
0.43	60.02	54.58
0.44	50.84	45.20
0.46	50.39	41.82
0.47	42.45	36.70
0.49	38.33	33.29
0.50	32.04	26.06

Interpreting these diagrams we derive the conclusion that both CAC_A and CAC_B outperform the reference scenarios, formulating a hybrid CAC algorithm of high performance. Moreover, emphasis must be laid on the fact that the convergence between the analytical (CAC_A) and the measurement-based (CAC_B) approach prove the accuracy of both methods. At the same time, the improved behaviour of CAC_HY in comparison with CAC_A and CAC_B justifies the need for their parallel implementation.

6 Conclusion

The presented hybrid CAC algorithm has been designed with the view of achieving the utmost scalability and adaptivity against the time-varying system parameters and propagation environment, so as to be able to support the envisioned versatility of the M-WiMAX. In this respect, it incorporates two autonomous modules for accurately calculating the upcoming bandwidth requirements of the connections under service. The first method performs a statistical analysis of the signal attenuation, while the second one makes virtue of physical layer info via a heuristic SNR sampling procedure. Finally, the hybrid CAC mechanism amalgamates these two supplementary approaches in order to determine the optimum admittance/rejection decision in respect to both the network's revenue and the satisfaction of the users' QoS.

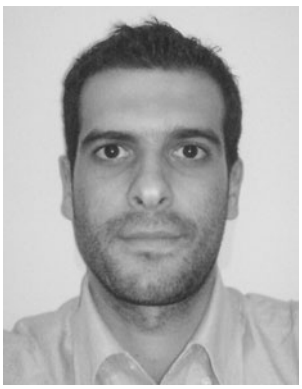
As it becomes evident from the results presented in Sect. 5, in the case of the CAC_HY algorithm, the BP is severely degraded in comparison with the reference CAC schemes, i.e. CAC_HY manages to increment the average number of simultaneously serviced connections via maximizing the utilization of the network resources. At the same time, for the whole range of the simulation scenarios, the contractual QoS requirements of all the flows are guaranteed in the absolute percentage of 100%, since the MSF is proved to

be kept above the minimum limitation of 1. As a matter of fact, not only a significant safety margin is secured (5% even for the hardest traffic conditions) but the provided QoS reaches the levels of the A_{R2} and A_{R1} (Fig. 13). On the other hand, the augmentation of the buffer size that is necessary in order to store larger amount of physical layer info ($N_{smp} - N_{R2,smp} = 7$) can be regarded as a fair price to pay for maximizing the utilization of the scarce wireless spectrum, especially in the framework of the contemporary vast capabilities of digital storage and processing.

References

1. IEEE Standard 802.16-2004. (2004). Air interface for fixed broadband wireless access systems.
2. IEEE Standard 802.16e-2005. (2005). Air interface for fixed and mobile broadband wireless access systems. Amendment for physical and MAC layers for combined fixed and mobile operation in licensed bands.
3. WiMAX Forum. (2006). Mobile WiMAX - Part I: A technical overview and performance evaluation.
4. Chakchai, S. I., Jain, R., Tamimi, A. K. (2009). Scheduling in IEEE 802.16e mobile WiMAX networks: Key issues and a survey. IEEE Journal on Selected Areas in Communications, 27, 156–171.
5. Niyato, D., & Hossain, E. (2007). Radio resource management games in wireless networks: An approach to bandwidth allocation and admission control for polling service in IEEE 802.16. *IEEE Wireless Communications, 14*, 27–35.
6. Niyato, D., & Hossain, E. (2005). Connection admission control algorithms for OFDM wireless networks. In *Proceedings of the IEEE global communications conference 2005 (GLOBECOM '05)*, vol. 5, p. 2459, St. Louis, Missouri, USA, December 2005.
7. Rong, B., Qian, Y., & Lu, K. (2007). Integrated downlink resource management for multiservice WiMAX networks. *IEEE Transactions on Mobile Computing, 6*, 621–632.
8. Elayoubi, S. E., Fourestie, B. (2008). Performance evaluation of admission control and adaptive modulation in OFDMA WiMax systems. *IEEE/ACM Transactions on Networking, 16*, 1200–1211.
9. Qin, C., Yu, G., Zhang, Z., Jia, H., & Huang, A. (2007). Power reservation-based admission control scheme for IEEE 802.16e OFDMA systems. In *Proceedings of the IEEE wireless communications and networking conference 2007 (WCNC '07)*, pp. 1831–1835, Hong Kong, March 2007.
10. Lee, J. Y., & Kim, K. B. (2008). Statistical connection admission control for mobile WiMAX systems. In *Proceedings of the IEEE wireless communications and networking Conference 2008 (WCNC '08)*, pp. 2003–2008, Las Vegas, Nevada, USA, March 2008.
11. Tsang, K. F., Lee, L. T., Tung, H. Y., Lam, R., Sun, Y. T., & Ko. K. T. (2010). Admission control scheme for mobile WiMAX networks. In *Proceedings of the IEEE International Symposium on Consumer Electronics 2007 (ISCE '07)*, pages 1–5, Dallas, TX, USA, 2007.
12. Kalikivayi, S., Misra, I. S., & Saha, K. (2008). Bandwidth and delay guaranteed call admission control scheme for QOS provisioning in IEEE 802.16e mobile WiMAX. In *Proceedings of the IEEE global telecommunications conference 2008 (GLOBECOM '08)*, pp. 1–6, New Orleans, LA, USA, 2008.
13. Zonoozi, M. M., & Dassanayake, P. (1997). User mobility modeling and characterization of mobility patterns. *IEEE Journal on Selected Areas in Communications, 15*(7), 1239–1252.
14. Liang, B., & Haas, Z. J. (2003). Predictive distance-based mobility management for multidimensional PCS networks. *IEEE/ACM Transactions on Networking, 11*(5), 718–732.
15. Soh, W. S., & Kim, H. S. (2006). A predictive bandwidth reservation scheme using mobile positioning and road topology information. *IEEE/ACM Transactions on Networking, 14*(5), 1078–1091.
16. Samaan, N., & Karmouch, A. (2005). A mobility prediction architecture based on contextual knowledge and spatial conceptual maps. *IEEE Transactions on Mobile Computing, 4*(6): 537–551.
17. Ma, W., Fang, Y., Lin, P. (2007). Mobility management strategy based on user mobility patterns in wireless networks. *IEEE Transactions on Vehicular Technology, 56*(1):322–330.
18. Ahson, S. A., & Ilyas, M. (Eds.) (2007) *WiMAX: Standards and security (Wimax handbook)*, 1 ed. Boca Raton: CRC Press.
19. Niyato, D., Hossain, E., & Diamond, J. (2007). IEEE 802.16/WiMAX-based broadband wireless access and its application for telemedicine/e-health services. *IEEE Wireless Communications, 14*, 72–83.
20. Ojanpera, T., & Prasad, R. (Eds.) (2001). *Wideband CDMA for third generation mobile communications*, 1 ed. London: Artech House Publishers.
21. Song, G., & Li, Y. (2005). Cross-layer optimization for OFDM wireless networks—Part II: Algorithm development. *IEEE Transactions on Wireless Communications, 4*(2), 625–634.
22. Goldsmith, A. (2005). *Wireless communications*, 1 ed. Cambridge: Cambridge University Press.
23. Simon, M. K., & Alouini, M. S. (2004). *Digital communication over fading channels*, 2 ed. New York: Wiley-IEEE Press.
24. Ahmed, M. H. (2005). Call admission control in wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 7*(1):49–68.
25. Theodoridis, G., & Pavlidou, F.-N. (2010). An SNR-based CAC algorithm for optimizing resource assignment in the downlink of M-WiMAX. In *Proceedings of the IEEE Wireless Communications & Networking Conference 2010 (WCNC '10)*, Sydney, Australia.

Author Biographies



Georgios Theodoridis received his diploma in electrical and computer engineering from Aristotle University of Thessaloniki, Greece, in 2004. He is currently working towards his Ph.D. degree in the same department. His research interests are in the field of call admission control and radio resource management in wireless terrestrial and Satellite/HAP networks. He is involved in Greek and European projects in these fields. He is a member of the Technical Chamber of Greece.

Fotini-Niovi Pavlidou received a Ph.D. degree in electrical engineering from Aristotle University of Thessaloniki, Greece, in 1988 and a diploma in mechanical-electrical engineering in 1979 from the same institution. She is currently a full professor at the Department of Electrical and Computer Engineering at Aristotle University, teaching in the undergraduate and post-graduate program in the areas of mobile communications and telecommunications networks. Her research interests are in the field of mobile and personal communications, satellite and HAP communications, multiple access systems, routing and traffic flow in networks, and QoS studies for multimedia applications over the Internet. She is involved in many national and international projects in these areas, and chaired the European COST262 Action on Spread Spectrum Techniques. She has served as a member of the TPC of many IEEE/IEE conferences. She is a permanent reviewer for many international journals. She has published about 80 papers in refereed journals and conferences. She is currently chairing the joint IEEE VTS & AESS Chapter in Greece.