# Admission Control in Mobile Cellular Networks

## Design, performance evaluation and analysis

Natalia Vassileva

TESI DOCTORAL UPC 2011

DIRECTORS DE LA TESI

Dr. Francisco Barceló-Arroyo
Dr. Yevgeni Koucheryavy

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Departament d'Enginyeria Telemàtica

Author

Natalia Vassileva
Department of Telematics Engineering
Technical University of Catalonia (UPC)
Spain

Supervisors

Assoc. Prof. Francisco Barceló
Department of Telematics Engineering
Technical University of Catalonia (UPC)
Spain

Prof. Yevgeni Koucheryavy
Department of Communication Engineering
Tampere University of Technology (TTY)
Finland

Reviewers

Assoc. Prof. Seferin Mirtchev
Department of Communication Networks
Technical University of Sofia (TU-Sofia)
Bulgaria

Adj. Prof. Giovanni Giambenne
Department of Information Engineering
University of Siena
Italy

Monograph set in LaTeX

*Perez de Ayala has stated it very wisely and skillfully: "Look at things as if for the very first time." That is, admire them afresh, disregarding what we remember from books, stilted descriptions, and conventional wisdom.*

Santiago Ramón y Cajal

*Advice for A Young Investigator*

*To my sister and parents*

# Abstract

Key to the success of the mobile telecommunication systems is provision of mobility and more specifically, continuity of communication sessions on the move. However, from engineering point of view the combination of scarce radio resource capacity, radio channel randomness, cellular structure, and user mobility can lead to call interruptions. The latter are regarded as very undesirable by mobile users. One essential part of the solution to this problem is to address resource insufficiency. To this end admission control algorithms are incorporated into the system. In this thesis we concentrate on admission control as a means of guaranteeing uninterrupted service to users with active calls on the move.

The research work reported in the monograph can be briefly summarised as follows. We explored the extensive empirical, analytical, and simulation results concerning teletraffic random variables of mobile cellular systems from the perspective of admission control. We proposed a conceptually different from prior work admission control solution based on the scientific evidence about the statistical nature of system variables and on the main result of renewal theory namely, decision making based on estimated system behaviour. In particular, we proposed a new admission control metric that uses statistical estimates. We evaluated the performance of the devised admission control strategy, which we named MRT (Mean Remaining Time) after the admission control condition, using both analytical and simulation approaches. We mathematically modelled system performance for traditional exponential conditions through a Markov chain. To study the MRT performance for non-conventional teletraffic scenarios we developed a simulation pure performance model. We examined the MRT for conditions that matched measured data from real, live mobile cellular networks. The results show that the scheme can guarantee call continuity and that it achieves a continuous working interval in contrast to the discrete one of the common cut-off scheme, yet the MRT strategy meets the important practical requirement for simplicity. We proposed an approximation to the MRT strategy for the case when not all of the required statistical information is readily available. The overall contribution of the aforedescribed research is that it is the first investigation that shows that statistical profiles of system random variables can be implemented into admission control strategies.

Next, we studied in-depth the implications of novel techniques introduced in advanced mobile telecommunication systems. In particular, we examined the effect of the adaptive modulation and coding (AMC) technique on system performance and consequently on admission control design in mobile WiMAX. The dynamic tuning to time-varying radio link conditions introduces new random variables that drastically change the traditional mobile cellular system model. In particular, cell capacity and call resource demands are not constant but random, determined by wireless link quality. We analytically modelled the radio channel randomness and the consequent non-deterministic resource demand for a streaming service with constant bit rate and strict delay requirements through a zone-based cell model. Furthermore, we examined system-level fairness, which metric had not been explored in previous studies on mobile WiMAX. Additionally, we studied the effect of AMC on system performance under the two basic admission control approaches proposed in the literature by incorporating them in the analytical model. The results show that the total new call blocking probability and forced call termination probability of a constant bit rate calls deteriorate when the radio channel conditions are quickly varying and the offered load is moderate to heavy. Furthermore, the results clearly show important differences in blocking and dropping probabilities of calls belonging to the same service (voice) and call (either new or handoff) class but being served in different modulation and coding zones. The results also indicate that if the admission control is not adapted to the actual system environment (dynamic in contrast to deterministic) it can lead to worse system performance compared to system performance when no admission control strategy is used.

# Preface

The research work presented in this thesis was carried out between the fall of 2006 and the fall of 2008 at the Department of Telematics Engineering (ENTEL), Technical University of Catalonia (UPC), Barcelona, Spain and between the fall of 2008 and the fall of 2010 at the Department of Communications Engineering (DCE), Tampere University of Technology (TUT), Tampere, Finland. Parts of the thesis were written during the same period but the main composition of the monograph took place from the end of 2010 until the spring of 2011 in Barcelona, Spain. Two months later, at the end of May 2011, the UPC supervisor shared his comments on the first draft with the author, whereas the TUT supervisor provided the author with his remarks on the second version before the end of July. At that time both supervisors authorised the thesis for submission to the Doctoral Committe at the Department of Telematics Engineering, UPC. The committee's approval came at the end of September, whereas the external reviews were made known to the author in mid-October. Administrative procedures relevant to the panel of examiners and documentation preparation preceeding the thesis deposit were carried out during the following month.

This thesis attained its present state due to the help of many people whose effort I would like to acknowledge.

I thank my supervisors Assoc. Prof. Francisco Barceló, ENTEL, UPC and Prof. Yevgeni Koucheryavy, DCE, TUT for their support, guidance and productive discussions. I am also thankful to Prof. Yevgeni Koucheryavy for fruitful and helpful discussions with experts from the telecommunication industry and academia in Finland and for his hospitality during my stay at Tampere. Also, I am grateful to Prof. Jarmo Harju, the Head of the Networks and Protocols Group (NPG) at DCE, TUT for having the door of his office open whenever I needed his wise advice and considerate support.

I am thankful to the thesis reviewers Adj. Prof. Giovani Giambenne, Department of Information Engineering, University of Sienna, Italy and Assoc. Prof. Seferin Mirtchev, Department of Communication Networks, TU-Sofia, Bulgaria for pre-examination and constructive feedback on my thesis. Moreover, I am grateful to Assoc. Prof. Seferin Mirtchev for initiating me into the art of scientific research at TU-Sofia. In addition, I would like to extend my thanks to all the professionals that contributed to my academic training and engineering education during the years.

I thank Dr. Israel-Martín, UPC for programming assistance with the development of the simulation tool for the first part of the thesis and to Dr. Dimitri Moltchanov, TUT for many interesting discussions the tangible outcome of which is a common publication.

I would like to thank my colleagues from UPC and TUT with whom I maintained a close contact for the pleasant working environment and many informal conversations during lunch and tea time. Particularly, I would like to mention: Dr. Marisa Catalán, Dr. Miquel Pericaz, M.Sc. Jacobo Sanz, M.Sc. Jordi Vilaseca; my colleagues from NPG at DCE, TUT and especially M.Sc. Jakub Jakubiak, M.Sc. Piotr Mitoraj, Dr. Roman Dunaytsev, Dr. Seppo Heikkinen, M.Sc. Tuure Vartiainen, M.Sc. Sari Kinari, Dr. Bilhanan Silverajan, and Dr. Jani Pietotalo. In addition, I am much obliged to M.Sc. Aleksi Suhonen for sharing his office at TUT with me and for fostering atmosphere where ideas and plants could flourish. Kiitos!

Grateful acknowledgement is made to Rosa Sanchez as well as to the previous secretaries at ENTEL, and especially for the help and diligence of Yolanda Forcada; to M.Sc. Jani Tuomisto, M.Sc. Tuure Vartiainen, M.Sc. Matti Tiainen as well as Sari Kinari and Ulla Siltaloppi from TUT for their prompt and expert technical and administrative aid.

My profound gratitude to Dr. Marko Kataja, Dr. Petteri Ingalsuo, Dr. Maija Kalliomäki and Arija is for their invaluable help, professionalism, and humanity; to Dr. Vladko Poposki and Dr. Dani Poposka for their unconditional support and care.

My warmest and most heartfelt thank you to my parents Ivanka and Vasil and to my sister Iva for their love, constant help, unselfish care, wise guidance and support throughout all my life as well as for giving me the right perspective to life's essential values. I thank my grantparents as well

as all my relatives and dear friends for their love and care.

The research work was financially supported by Ministerio de Asuntos Exteriores y de Cooperación–Agencia Española para Cooperación Internacional (MEC-AECI), Ministerio de Educación y Ciencia (Beca Predoctoral de Formación de Personal Investigador, FPI), Ministerio de Ciencia y Tecnologa, MCYT (under the projects TEC2006-09466/TCM: "Localización de terminales en redes celulares y sus aplicaciones a la optimización de la operación de la red" and TIC2003-01748: "Red ubicua basada en IP (RUBI): anlisis y desarrollo de un prototipo para optimización de recursos con localización de terminales") and the Finnish Government (through the Centre for International Mobility Fellowship), all of which are greatly acknowledged.

Finally, I consider myself fortunate for having as additional sources of motivation for my work the cosmopolitan Barcelona with its unique architecture, luxurious climate, rich cultural programme, Mediterranean cuisine, football (Visca Barça!) and smilling people as well as the inspiring beauty of the incredible nature of Finland, the avant-garde Finnish style, genuine people, educational values and philosophy of life.

Autumn 2011, Barcelona
*Natalia Vassileva*

# Contents

# List of Figures

# List of Tables

# 1 Introduction

The concept of the mobile communications—freedom to communicate untethered, on the move, and without restrictions on place or time—has profoundly caught on with virtually every person regardless of their race, whereabouts, educational background, economic status, or age. The number of mobile subscribers has surpassed the number of fixed Internet connections. What is more there are people who do not have electricity at home but do have mobile phones [3]. The mobile cellular systems have become ubiquitous. The offered services are constantly diversified (mobile video, location-based services, etc.); new mobile devices (smartphones, tablets, laptops) and new communication paradigms (machine-to-machine) are introduced throughout the years.

The increased capacity demands of the enhanced services compared to the plain voice service, the intensified usage of the communication tools, the heavier traffic load generated by the smart devices compared to the one typical for the ordinary mobile phones as well as the traffic migration to the mobile network from the wired network are contributing to the continuous growth of the mobile traffic. Thereby, despite the significant advances in the development of the wireless technologies, which have made possible the better exploitation of the scarce radio spectrum, the capacity in the access part of the wireless networks remains deficient.

Although mobility and more specifically, continuity of communication sessions on the move is key to the success of the mobile telecommunication systems, from engineering point of view mobility support is a challenging task. The combination of scarce radio resource capacity, radio channel randomness, cellular structure, and user mobility can lead to call interruptions. The latter are regarded as highly undesirable by mobile users. One essential part of the solution to this problem is to address resource insufficiency. To this end admission control algorithms are incorporated into the system. In this thesis we concentrate on admission control as a means of guaranteeing uninterrupted service to users with active calls on the move.

There are two general scenarios of practical interest for which system resource insufficiency can result in communication session interruption and can prevent mobile wireless networks from meeting agreed quality of service. Firstly, a communication session can be initiated and terminated in different cells of a network because of users' mobility and network's cellular structure. If one of the cells that a mobile user visits for the communication session duration lacks enough resources to serve the ongoing session, the latter will be forced to terminate (see Fig. 1.1). Secondly, the amount of resources needed to serve a communication session with agreed quality parameters generally depends on the radio link conditions, which are randomly changing in time due to the very nature of the physical medium used for transmission. Often, when the signal quality of the radio channel between the base station and mobile station deteriorates, additional resources must be allocated to the ongoing session in order to maintain its quality of service (QoS) level or (and) its successful continuation (see Fig. 1.2).

The objective of admission control (AC) in mobile cellular systems is to efficiently manage the system resources so that the quality of service of the admitted communication sessions can be guaranteed until their voluntary finalisation ragardless of user mobility or experienced radio link conditions. An important practical requirement for admission control is simplicity [31, 111] because the admission control is executed on call request arrival (that is, frequently), must give a quick acceptance (rejection) response and shall be easy to implement. In practice, wireless network operators choose admission control schemes based on the level of the algorithm's practical incorporation complexity and overall performance.

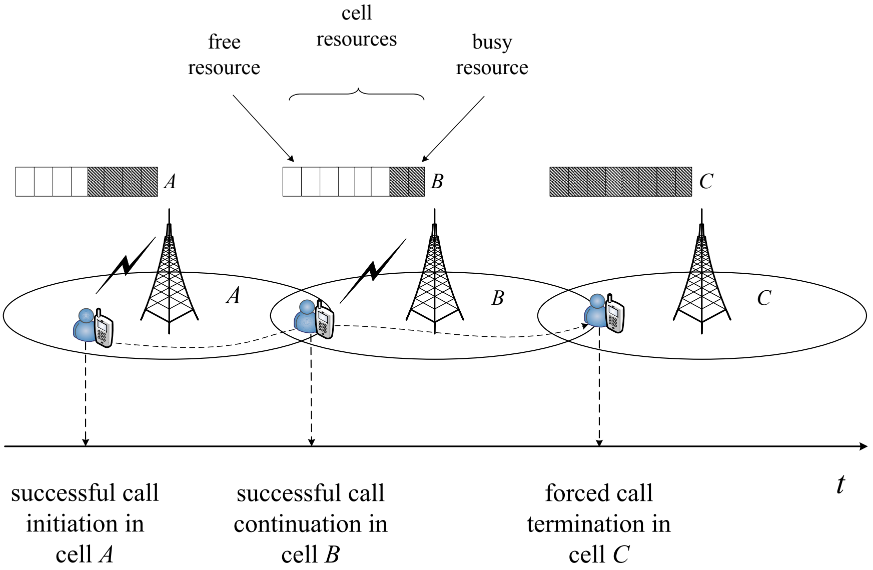

cell
resources
free
resource
busy
resource
A
B
C
A
B
C
t
successful call initiation in cell A
successful call continuation in cell B
forced call termination in cell C

Figure 1.1: Forced call termination due to insufficient resources at the target cell $C$ during inter-cell handoff

## Contributions

The research work reported in this monograph has been aimed at contributing to the performance evaluation, analysis and design of admission control algorithms in mobile cellular networks.

In Part I of the thesis we focus on the design of admission control solution that adresses the scenario depicted in Fig. 1.1[1]. The problem has been a focus of research since the development of the first mobile communi-

[1] The earlier generations of mobile wireless networks were designed for the worst case radio conditions experienced at the far edge of the cell. The capacity of these systems as well as the allocation of resources to accepted calls are fixed. We assume this conventional model in the first part of our investigations. Later, in Part II, we consider dynamic capacity and resource allocation conditions, which are typical for the modern wireless communication technologies.

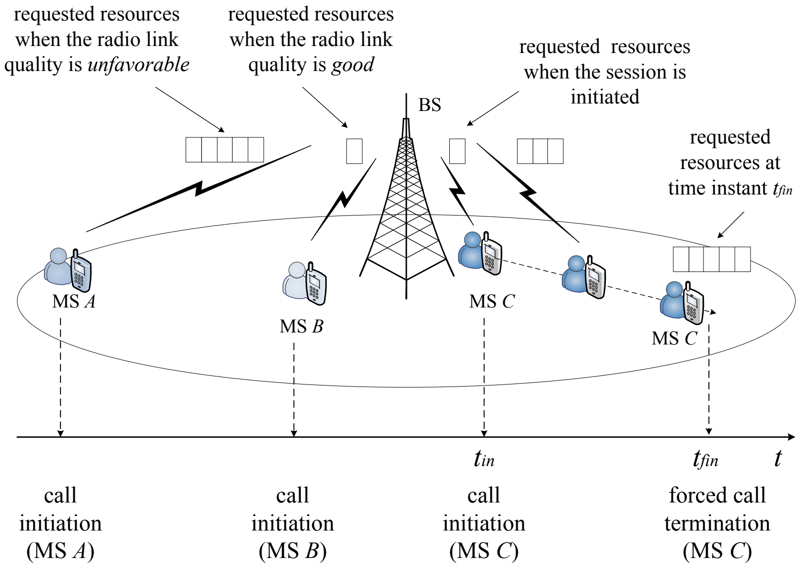


Figure 1.2: Forced call termination due to insufficient resources at the serving cell when the radio link conditions between the base station (BS) and mobile station (MS) deteriorate

cation systems and valuable solutions to it have been proposed in the past. However, one potentially feasible and efficient approach was not explored previously, namely the application of the teletraffic random variables' profile in improving system performance. In effect, there has been considerable research effort devoted to the probabilistic characterisation of teletraffic variables that describe mobile cellular systems. Despite the fact that these statistical investigations (which are to be overviewed in Chapter 2) had been motivated by system performance objectives, their outcomes were not considered for devising admission control solutions previously.

In the first part of the monograph we explore the statistical profile of relevant teletraffic random variables from the perspective of admission control. Importantly, we demonstrate that the statistical properties of these variables can be advantageously used in the development of admission control strategies that enhance system performance and user satisfaction.

Part II of the thesis is devoted to the anaysis and performance evaluation of advanced mobile telecommunication systems. Unlike the conventional mobile systems, for which the allocation of resources to communication sessions is fixed, the broadband mobile systems are characterised with non-deterministic cell capacity and session resource demands (see Fig. 1.2). The latter two are main metrics in admission control. It is of practical interest therefore, to assess the new context within which the modern systems perform and admission control is executed.

In the second part of the thesis we investigate the effect of the adaptive modulation and coding (AMC) technique on system performance and its repercussion on admission control design. We show that the dynamic resource conditions can yield system-level unfairness and negatively impact those connections with strict (delay, bit rate) QoS requirements that experience unfavourable radio conditions. Importantly, we show that the resource environment for systems that implement AMC can not be assumed fixed and that the effect of the non-deterministic environment must be addressed by admission control.

### Overview of the monograph

The monograph is organised in eight chapters including the present one. Having succinctly stated the problem and motivation for our research, in Chapter 2 following we continue with a brief introduction to the basic system model of land mobile cellular networks. The metrcis commonly used for mobile system performance evaluation are overviewed. Pertinent teletraffic random variables are described and investigations dedicated to their statistical modelling are concisely overviewed.

The reminder of the monograph is structured in two parts. The first part concerns the traditional mobile cellular networks and comprises chapters 3 to 5. The second part explores research questions relevant to the modern, advanced mobile telecommunication systems and consists of chapters 6 and 7. Each chapter is initiated with a short introduction that explains the particular motivation and goals as well as the contributions of the research reported in the chapter.

In Chapter 3 we devise a conceptually different admission control scheme that builds on the scientific evidence about the statistical properties of the teletraffic variables of interest and the fundamental concept exploited in reliability theory and practice, namely decision making based on the estimated future system behaviour. The developed scheme attains a larger working interval than the classical cut-off scheme, which is advantageous because a wider working interval provides mobile operators with more freedom to make a trade-off between quality of service and efficient resource use.

In Chapter 4 we complement the scheme's assessment by evaluating it for conditions typical for the present and future land mobile wireless systems, namely high handoff rates, overlapping areas, and non-Poisson arrival traffic flows. Relevant for the practice conclusions concerning the mobility degradation problem and admission control tuning are drawn from that study.

Lastly, in Part I, Chapter 5 the functionality of the teletraffic-based admission control scheme is examined for conditions more restrictive than those assumed in its design. For such conditions we suggest an approximation of the main parameter of the devised in Chapter 3 admission control scheme. The investigation results reported in Chapter 5 lend validity to the proposed approximation approach but also reveal some scheme limitations.

A succinct overview of the main buliding blocks of the advanced wireless cellular systems is provided in Chapter 6. These are shortly explained, primarily from the angle of the research objectives set in Part II of the monograph. We also provide a concise summary of the WiMAX technology, which was considered in our investigations because mobile WiMAX is representative for the majority of the foremost broadband standards.

The effect of the adaptive modulation and coding technique on system-level performance is studied in Chapter 7 through mathematical anaysis. The effect of the random radio link behaviour on the frequency with which an ongoing session is interrupted due to insufficient resources in the cell as well as the lack of fair service are analysed. General guidelines for admission control devising are suggested.

Finally, in Chapter 8 a recapitulation of the reported research work and the main conclusions drawn from it is done and possible future research lines from the perspective of the work carried out are outlined.

# 2 System model and parameters

This chapter is devoted to the basic characteristics of mobile cellular networks. The system model (see Section 2.1) and performance metrics used for system evaluation (see Section 2.2) are described. Special attention is paid to the teletraffic parameters pertinent to these networks (see Section 2.3). The main goal is to overview in a succinct yet sufficiently detailed form the most important results of the extensive research concerned with statistical modelling of teletraffic random variables relevant to mobile cellular networks. These results serve as a basis for the research reported in the remainder of the thesis. Lastly, we conclude on the overviewed results (see Section 2.4).

## 2.1 Model description

The fundamental characteristics of a land mobile system with cellular structure comprise the following:

- The network area is divided into geographically distinct but contiguous areas called *cells* (see Fig. 2.1) each controlled by a *base sta-*

*tion* (BS)[1]. The connection between a *mobile station* (MS) and a BS is accomplished via a radio link.

- A MS can be in one of two states: active or inactive. The time interval during which the terminal is in active state (communicating) and is using allocated resources is denominated *call*[2].

- The MSs, as the name suggests, are *mobile*, therefore users can change their initial, with respect to the instant when the call was initiated, location and traverse several cells for the call duration. Two fundamental mechanisms for supporting continuous service to *mobile users* (MUs) with ongoing calls are: handoff mechanism and overlapping cell areas.

- When a MU engaged in a call crosses the cell boundaries of the serving cell, a handoff is triggered. *Handoff* (or *handover*) is the process of migration of a mobile terminal with a call in progress form the air interface of a serving BS to the air interface of a target BS; the radio resources assigned to the call in the serving BS are released, whereas new radio resources are requested from the target BS.

- The radio coverage areas of neighbouring cells overlap to enable uninterrupted service to the mobile users. The portions of the network where service is provided by more than one BS are called *overlapping* (or *handoff*) areas.

In admission control system modelling it is generally assumed that the underlying handoff process is ideal and handoff call requests are dropped only due to resource insufficiency (see Section 2.2), which was the approach used in our work as well.

[1] The *cellular concept* (the division of the area covered by a mobile wireless network into distinct geographical clusters) is introduced to allow for reutilisation of the radio spectrum thus, for increased number of simultaneously served users. The cellular structure therefore, achieves larger carried traffic per unit area; that is, it improves the communication capacity.

[2] For the sake of conciseness we use the term *call* instead of the more general *communication session.*

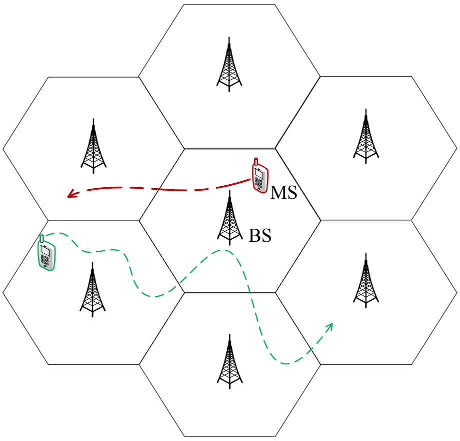


Figure 2.1: Layout of a mobile cellular network

## 2.2 Performance metrics

Mobile cellular system-level performance is evaluated through forced call termination and new call blocking probabilities (or alternatively to these two, call completion probability) and carried traffic. The first two metrics measure the quality of service provided by the system to the users, whereas the carried traffic measures the efficiency of system resource use. These performance metrics serve also for assessing the improvements in system performance introduced by admission control strategies.

*New call blocking probability*, denoted $Pb$, is the probability that a new call request will be rejected by the system. Likewise, *dropping probability*, denoted $Pd$, is the probability that an ongoing call request will be dropped by the system due to insufficeint resources. *Forced call termination probability*, $Pft$, is the probability that a call admitted into the system will be forced to terminate at some point during its course (i.e., the call will not be finished voluntary by the user but terminated by the system) due to insufficient resources. It is noteworthy that dropping probability is a measure of the frequency with which ongoing calls are rejected at a base station, whereas forced call termination probability measures the frequency with wich ongoing calls are interrupted. The International Telecommunication Union – Telecommunication Standartization Sector (ITU-T) stipulates[3] that the dropping probability of ongoing calls should be (orders of magnitude) lower than the blocking probability of new calls. This guideline is followed in the open literature because it is commonly accepted that the mobile user perceives call interruption as more disturbing than call initiation rejection. Therefore, admission control algorithms prioritise handoff (or *active*, *ongoing*) call requests in front of new call requests.

More recent studies evaluate system performance considering the unreliability of the radio channel in addition to resource insufficiency (see [4, 73, 82, 115–117]). Despite the practical interest in such a scenario, it is beyond the scope of the work reported here. The main reason is that forced call termination due to severally degraded radio link between a BS and a MS (i.e., physical link break-down) is not under the control of admission control designers; that is, the factors that govern the state of the radio link can not be readily controlled by AC algorithms. Furthermore, according to the ITU-T Recommendation E.771 [1], the probability of unsuccessful land cellular handover: " $\cdots$ is the probability that a handover attempt fails because of lack of radio resources in the target cell, or because of a lack of free resources for establishing the new network connection", implicitly suggesting that the handover process is assumed to be reliable (i.e., a link with a base station is maintained). Therefore, we adopt the common assumptions for ideal underlying handoff process and for forced call termination due only to lack of radio resources. In particular, in Part I of the thesis we focus on conventional mobile wireless systems, which typically have a fixed link margin. Hence, for such conditions and the assumption for reliable radio link, forced call interruption can take place only during a handoff to a neighbour cell with insufficient free resources. In Part II we centre on modern mobile wireless systems, which typically incorportate link-adaptation techniques. Therefore, we model the fluctuating radio channel but physical link breakdown is again not considered for the reasons discussed above. In mobile cellular systems with link-adaptation a call can be forced to terminate not

[3] See ITU-T Recommendation E.771 [1] that concerns performance metrics of land mobile networks

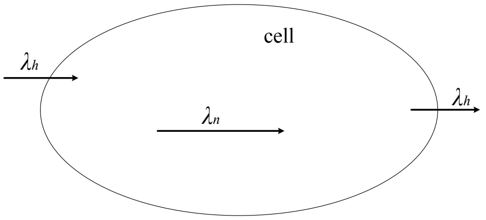


Figure 2.2: Arrival process components: (i) new call traffic flow with rate $\lambda_n$ and (ii) handoff call traffic flow with rate $\lambda_h$

only during handoff because of insufficient resources in the traget cell but also due to insufficent resources in the serving cell; namely, when the radio link is deteriorated and the additional resources that must be allocated to the call in order to maintain agreed quality of service are unavailable.

## 2.3 Teletraffic parameters

In this section we summarise relevant to mobile cellular systems teletraffic parameters and overview research dedicated to their probabilistic description.

### 2.3.1 Arrival process

Different from the arrival process in wired networks, which is generated only by stationary users, the arrival traffic in mobile cellular networks comprises: (i) *new* (*fresh*) calls that are originated by MSs located in a studied cell and (ii) *handoff* (*ongoing*, (*pre-*)*existing*, or *active*) calls that have been carried out in neighbouring cells before MSs move to and request service from the cell of interest (see Fig. 2.2). We look at these traffic flows as well as at the *aggregate* (*merged*) arrival traffic.

### New call arrival traffic

It has been widely accepted that in *fixed* (*land-line*) telephone networks: (i) the call arrivals are independent of each other and (ii) the interarrival times between call requests are exponentially distributed. Therefore, traditionally, the new call arrival traffic is modelled as a Poisson process[4]. This practice is commonly followed when modelling mobile cellular networks mainly because it is assumed that the subscriber population from which new calls originate is large and as a result can be approximated by an infinite population[5].

### Handoff call arrival traffic

Contrary to the extensively used and accepted hypothesis for Poisson new call arrival traffic, the nature of the handoff process is questioned in a significant number of publications. The latter is motivated by the fact that the mobile cellular networks from second generation (2G) on feature micro and pico cells in urban areas[6]. When the cell size is shrunk, the number of handoffs that a call goes through for its duration is increased. Thereby,

[4] Recall that a counting process $\{N(t), t \geq 0\}$ is said to be a *Poisson process* with rate $\lambda$ ($\lambda > 0$) if: (i) $N(0) = 0$, (ii) the processs has independent increments, and (iii) the number of events in any interval of lenght $t$ is Poisson distributed with mean $\lambda t$, that is: $P\{N(t+s) - N(s) = n\} = e^{-\lambda t}\lambda t^n/n!$, where $n = 0, 1, \ldots$, and $s$, $t \geq 0$. It follows from condition (iii) that a Poisson process has stationary increments and also that $E[N(t)] = \lambda t$ [83]. In short, a Poisson process is a counting process for which the times between successive events are independent and identically distributed exponential random variables [83].

[5] For systems with infinite population the number of subscribers served (number of ongoing calls) in the system does not have effect on the arrival traffic. Such systems are easier to describe and analyse. Thereby, when the population is large and the mean arrival rate is not altered by the number of ongoing calls the population is approximated by an infinite source.

[6] The main reason behind such an (small-size-cell) architecture is that it allows for dense reutilisation of the scarce radio spectrum and accommodation of a larger number of subscribers. When the area covered by a BS is small, the power requirements of mobile terminals are drastically decreased and consequently, the power emitted by the BS is considerably reduced. The resulting inter-cell interference is much lower compared to the case of large cell coverage areas.

assuming that on average a call requires several handoffs—that is, visits several base stations—it can be conjectured that the handoff arrival process (that is, the handoff traffic profile) diverges from the Poisson one and is smooth. Another motivation for the intensive research on the handoff arrival behaviour is the fact that the handoff traffic is generated by calls that are served by neighbouring with respect to the target cell BSs. The number of handoff sources therefore, is limited by the maximum number of ongoing calls in the immediate neighbourhood of the considered (target) cell and hence, can not be approximated by an infinite population. Therefore, several authors, as described below, investigate the validity of the exponential handoff arrival traffic hypothesis.

The probability distribution of the handoff traffic flow was examined first by Chlebus and Ludwin [20]. The authors demonstrate that the Poisson handoff assumption holds for the ideal case of infinite number of communication resources (call requests are never blocked), whereas for the realistic case of blocking environment with finite number of resources, the handoff traffic is *smooth*[7]. Nevertheless, after comparing the approximate analytical with simulation results, Chlebus and Ludwin conclude that for light and moderate traffic load the handoff arrival process is very close to a Poisson process and therefore, for such traffic conditions the Poisson hypothesis is a reasonable working assumption.

Sidi and Starobinski [89] argue that system models that consider the handoff traffic as a Poisson process are reasonable when the traffic is homogeneous and the cellular area is wide. Conversely, when the number of cells is small or the traffic is not homogeneous the authors conclude that the Poisson assumption is inaccurate.

In a number of publications Rajaratnam and Takawira (see [76–78] for instance) show through mathematical and simulation studies that indeed the offered handoff traffic is smooth. The authors study the handoff process under negative exponential channel holding time (see below), various mobility models and traffic loads [77]. Later, the same authors conduct a similar study but under general holding time (gamma and the proposed by them det-neg distributions) [78]. For the investigated probability distribu-

[7] A *smooth* process has a variance that is less than its mean, i.e., $Var[X] < E[X]$.

tions the numerical and simulation results suggest that the peakedness of the arrival traffic of ongoing calls is always less than one, which conclusion is analytically proven by van Doorn and Ta [105].

Zeng and Chlamtac [113] focus on the impact that different teletraffic and mobility parameters have on the interarrival handoff distribution and rate for both blocking and non-blocking conditions. The main conclusion that the authors derive form the simulation results with respect to the handoff pattern is that the handoff arrivals can only be approximated by a Poisson process when the load generated by ongoing calls is large and the environment is non-blocking.

Unlike previous research that focuses on a single cell, Orlik and Rappaport [69] model a cluster of seven cells and evaluate the performance of the central cell rather than just of the isolated cell of interest. The authors develop approximate analytical models for two cases with respect to the handoff traffic offered to the cluster: one that assumes Poisson process and another that uses a *Markov modulated Poisson process* (MMPP)[8]. Orlik and Rappaport reason that the differences between the results obtained through the proposed "isolated cluster" model and the "single isolated cell, Poisson handoff arrival" model are insignificant, especially under heavy traffic loading. The authors suggest using the traditional "single isolated cell, Poisson handoff arrival" model, first proposed by Rappaport [79], for planning and analytical purposes because of its simplicity.

Martin-Escalona *et al.* [67] develop a simulation tool that allows statistical characterisation of relevant teletraffic variables for more realistic conditions. They model a square-shaped cluster area with overlapping hexagonal cells (36 cells in total). The interarrival handoff statistics are collected for *log-normally* distributed unencumbered call duration[9], high-speed users (see Table 3 in [67]), and non-priority scheme (new and ongoing calls are treated equally). For these conditions and when 50 % or more of the traffic load is generated by handoff arrivals, the time between two consecutive handoff arrivals is characterised with $Var[X] < E[X]$ and best fitted by hyper-Erlang [67]. The simulator presented in [67] is extended further by

[8] MMPP is a Poisson process whose rate varies according to a Markov process.

[9] This is in accordance with measured data reported in empirical studies.

Spedalieri *et. al* [96] to model the particularities of a UMTS network. For the aforementioned conditions (log-normally distributed call duration, etc.) it is shown that the interarrival times between consecutive handoff requests can be modelled by *Erlang-k* (light offered traffic) or *gamma* (medium and heavy traffic load) distributions.

#### Aggregate arrival process

Li and Alfa [8, 64] analytically model the *aggregate* (*merged*) arrival traffic of new and handoff calls through *Markov arrival process* (MAP)[10] with correlated interarrival times, which is an exception from the general approach of assuming aggregate Poisson process.

The literature does not abound with empirical investigations on the cellular arrival traffic profile either. Jedrzycki and Leung [55] report on the channel occupancy time and call interarrival distributions measured in a real telephone mobile cellular network. They fit the call interarrival times with a negative exponential distribution (n.e.d.) [55]. Contrary to Jedrzycki and Leung, however, based on collected empirical data from a working mobile cellular network, Barceló and Sánchez [15] conclude that the aggregate arrival traffic is not Poisson but smooth. In particular, the authors find that the time between call arrivals is best fitted by *Erlang* distribution [15]. This same result was previously reached by Barceló and Bueno [12] for a *Public Access Mobile Radio* (PAMR) system.

### 2.3.2 Cell residence time

*Cell residence* (or *dwell*) *time* is the duration of time that a (mobile) user (regardless of whether involved in a communication session or not) spends in the vicinity of a BS (cell); see $\Delta t$, Fig. 2.3. It depends on factors such as cell shape and radius, user mobility and trajectory, etc. The majority of the admission control proposals (see [5, 31, 42, 88] and the surveyed literature therein) make a classical assumption of a negative exponential cell dwell time. The works that are an exception of the aforementioned approach

[10] MAP is a class of point processes. The MMPP is a special case of the MAP.

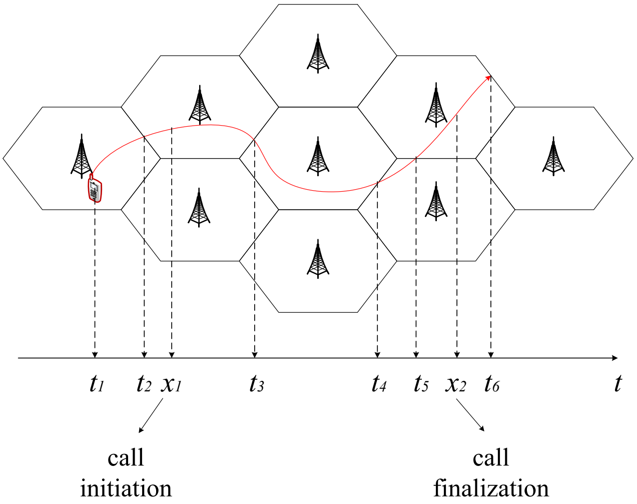


Figure 2.3: Teletraffic variables: (i) cell residence time ($\Delta t = t_{i+1} - t_i$); (ii) unencumbered call duration ($\Delta x = x_2 - x_1$); (iii) channel holding time ($t_3 - x_1$, $t_4 - t_3$, $t_5 - t_4$, $x_2 - t_5$).

model the cell residence time with general distributions: gamma [18, 32, 58, 118], sum of hyper-exponential distributions [58, 71], Erlang or hyper-Erlang [39, 58], and Weibull [58, 72].

In a series of articles Kobayashi *et. al* [47–49, 86] report on their empirical findings regarding the cell dwell time parameter. The authors [48, 86] demonstrate that from the examined exponential, Weibull, gamma and log-normal distributions the latter one is the closest approximation to the measured data. It is concluded that when the cell size is small the cell residence time depends on the vehicle motion, whereas when the cell size is large it depends on the call holding time. In addition, whereas the speed and estimated direction of movement diverge in small and large cities[11], the cell dwell time exhibits the same distribution. Furthermore, the authors discover a self-similarity in the taxi's cell dwell time characteristics [47, 49].

[11] The field measurements are carried out in Japan, so the notion of 'small' and 'large' shall be considered within that context.

### 2.3.3 Call occupancy time

The *unencumbered call duration* is the intended time duartion of requested call connection, which corresponds to the holding time in traditional wireline networks, where premature (forced) call termination and active call channel vacanting are not considered [71]. This is the $\Delta x$ time interval shown in Fig. 2.3 and is often called just *call holding* (or *occupancy*) *time.* Fang and Chlamtac differentiate between *actual* call connection time [39] (or *effective* call holding time [38]) and *requested* call connection (or session) time. The former refers to the duration of an actual call connection of an incomplete call and depends on the network (call interruption due to insufficient resources), whereas the latter is, in effect, the unencumbered call duration of a complete call and is determined by the mobile user (how long the user wants to maintain the call). We overview here only the empirical works that statistically describe the call occupancy time variable as in general the analytical and simulation studies assume rather than investigate its probability density distribution.

A recent field study of Yavuz and Leung [112] shows that not only the channel holding time is best approximated by a log-normal distribution but that the call holding time of both *mobile* and *stationary* users is of *log-normal* type as well. The authors stress that the call holding distribution radically differs from the exponential one. This result is consistent with statistical investigations on the unencumbered call holding time in fixed telephony: Bolotin [16] states that the mix of log-normal distributions is a closer fit to collected data than the classically used negative exponential distribution. Chlebus [19] indicates that the call duration follows the patterns reported by Bolotin [16] according to [14].

### 2.3.4 Channel holding time

In mobile cellular systems, due to user mobility and system cellular structure, the *resource holding* or *occupancy time* usually differs from the call holding time. Mobile users that have started their calls in one cell of the network may move to adjacent cells for the duration of their calls. The resource holding time is defined as the time interval from the instant cell resources are allocated to a call request (either new or handoff) until the instant they are released by the call due to: (i) successful (forced) termination of the communication session or (ii) continuation of the service into an adjacent cell. In short, the resource occupancy time is the time during which the call occupies communication resources within a cell during its residence in the cell (see Fig. 2.3). The resource holding time is commonly known as *channel holding time* (CHT) as the term was first introduced in the conventional voice-oriented mobile networks, which feature one service and a fixed resource allocation. *Channel*[12] denotes the resources allocated by a base station to an accepted call (either new or handoff). The channel holding time depends on factors such as user velocity, direction and initial position, cell size and shape, propagation conditions as well as on the call holding time variable.

The channel holding time distribution has been a focus of research for many authors. Hong and Rappaport [50], based on approximate *analytical study*, affirm that the negative exponential distribution could approximate the channel holding time distribution when the call hodling time also follows exponential distribution. Soong and Barria [95] explore the relationship between the cell residence time and handoff channel holding time and conclude that when the cell residence time is of Erlang type, the channel occupancy time distribution is equivalent to a Cox model. Rajaratnam and Takawira [77] argue that this is the case for large cell sizes, whereas for reduced cell sizes and gamma distributed cell residence time, the CHT is better modelled by probability distributions smoother than the exponential, such as the Erlang-*kones*. Wang and Fan [108] examine the occupancy ditsribution under general distributions of the unencumbered call holding and cell residence time. The authors show that the studied variable is indeed sensitive to the first and second moments of the call holding time. In particular, when both the call occupancy and cell residence time are of Erlang type, the analytical results diverge significantly from those obtained under the classical exponential assumption. The conclusions based on the work of Christensen *et al.* [21], who investigate the channel holding time distribution dependance on other teletraffic random variables, are in line

[12] The term *channel* refers to a single resource: a fixed frequency bandwidth, time slot, or code depending on the used access technique (FDMA, TDMA, or CDMA).

with the results of Wang and Fan [108]. Specifically, Christensen *et al.* show that the channel holding time distribution is of phase type provided that the cell residence and call holding type can also be represented by phase type distributions [21]. In a series of publications Corral-Ruiz *et. al* [22–24] examine as well the relationship between the cell residence time and the expected channel holding time. The authors demonstrate that the channel holding time has the same distribution as the cell residence time (Coxian, hyper-exponential, or hyper-Erlang).

It should be pointed out that the overviewed mathematical investigations on the CHT show that when the cell residence time or (and) the call holding time are analytically modelled with a probability distribution function other than the exponential one, the channel holding time (resource occupancy time) shows behaviour that is far from *memoryless.*

The *simulation study* carried out by Guerin [45] demonstrates that the channel occupancy time distribution might display a rather poor agreement with the n.e.d., especially for mobile users with low changes in rate of movement and direction. More versatile *simulation* tool for determining the probability distribution of key teletraffic parameters is developed by Zonoozi and Dassanayake [118]. The authors examine the channel holding time distribution for exponential unencumbered call holding time (as it was done by Hong and Rappaport [50], and Guerin [45]) and gamma distributed cell residence time. The results obtained under these conditions support the n.e.d. channel holding time distribution hypothesis. Khan and Zeglache [58] find that the resource holding time is n.e.d. when both the cell residence time and the call holding time are exponentially distributed. Spedalieri *et. al* [96] show through simulation experiments that the CHT is best fitted through hyper-Erlang-$jk$ distribution. The experiments are carried out under log-normally distributed call duration. The latter simulation results [58, 96] are in agreement with the analytical works that show that when the cell residence and call holding time are non-exponentially distributed, the channel occupancy time is non-memoryless as well.

The first *field study* that determines which probability distribution fits best the voice channel holding time in a mobile cellular network is reported by Jedrzycki and Leung [55]. The authors observe that the log-normal distribution fits the channel occupancy time data much better than the memoryless distribution. A similar observation—the exponential distribution shows a poor fit with measured statistics—is reported by Barceló and Jordán [13, 14]. The authors carry out independent field trials and state that the log-normal or a mixture of log-normal distributions (in particular, log-normal-3 [13]) suits better the collected data for the channel occupancy time than the exponential distribution [13, 14]. The empirical research of Hidaka *et. al* [49] points out that mobile cellular systems with predominant data traffic and micro-cells are more likely to evidence long-tailed log-normally distributed cell residence times and thereby channel holding time that exhibits self-similarity. The same auhtors add that channel occupancy time is affected to a greater extend by the cell rather than the call holding time when the latter is long. More recent statistical data from field measurements in a live mobile cellular system reported by Yavuz and Leung [112] confirms the results of the previous empirical studies: the negative exponential distribution is a poor approximation of actual data. In particular, Yavuz and Leung [112] show that the log-normal distribution is the best approximation for all classes of channel occupancy times. It is important to note that the overviewed field studies [13, 14, 55, 112] follow a common approach: the observed real statistics are first compared with the exponential distribution and only then a better fit is searched.

The main conclusion that can be drawn from the reported empirical results is that distributions other than the negative exponential one are closer to measured, real channel occupancy data.

Motivated by the observed statistcis in functional mobile cellular networks (in Canada [55,112] and Barcelona [13,14][13]) several researchers focus on finding a versatile probability distribution that can concurrently suit *observed data* and *mathematical analysis* tractability. Orlik and Rappaport propose a sum of hyper-exponential (SOHYP) distribution to adjust the channel holding time variable to the reported empirical data. According to the same authors the SOHYP can be used to approximate exponential, hyper-exponential, and Erlang distributions (i.e., SOHYP can have

[13] The empirical studies carried out in Japan [47–49, 86] have received less attention perhaps due to limited access to the publications.

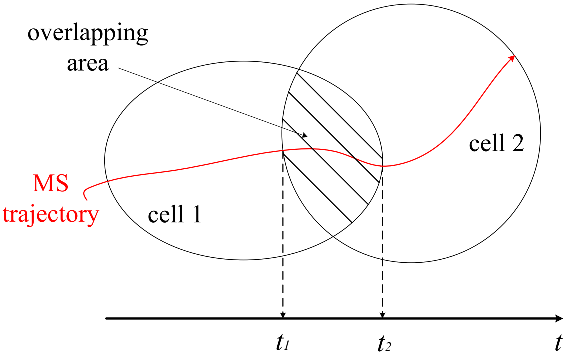


Figure 2.4: Residence time in the area covered by more than one base station (handoff dwell time): $\Delta t_r = t_2 - t_1$

*coefficient of variation* ($CV^{14}$) smaller, equal or greater then 1) Fang and Chlamtac propose a hyper-Erlang distribution for modelling the cell residence time so that the reported log-normal distribution of the CHT can be fitted [39]. Christensen *et al.* [21] suggest phase-type distributions for modeling the channel holding time as well as other teletraffic variables such as the cell dwell and call holding times. The framework they porpose [21] is flexible and versatile as it incorporates the proposed Cox, SOHYP, and hyper-Erlang models discussed previously.

### 2.3.5 Handoff dwell time

The overlapping (handoff) cell area is a portion of the network coverage area served by more than one base station as noted earlier. The *handoff dwell* (or *sojourn*) *time* is the time spent by the mobile user in the overlapping area as shown in Fig. 2.4, see $\Delta t_r$.

[14] $CV = \sigma/E[X]$, where $\sigma$ is the standard variation, $\sigma = \sqrt{Var[X]}$, $Var[X]$ is the variation, and $E[X]$ is the mean of the distribution.

The exponential distribution is widely used in the literature for modelling the handoff dwell time and only few works investigate the validity of this hypothesis (see [75,84] and references therein). The results reported by Pla and Casares-Giner [75] in the most recent of these publications suggest that the n.e.d. approximates well the handoff dwell time. Further, their numerical results [75] show that for low values of the variance of the handoff dwell time the use of exponential distribution yields more conservative results.

### 2.3.6 Comments on the reported analytical, simulation and empirical results concerning teletraffic variables

Studies that investigate the statistical nature of teletraffic variables pertinent to mobile cellular systems sometimes reach apparently different conclusions. It is noteworthy that the research works have often been conducted under different *modelling assumptions* and by adopting different *approaches*, which could explain seemingly discrepent results reported in the literature. Consider for instance the resource occupancy time parameter. It is demonstrated that the channel holding time depends on both cell residence and call holding time distributions [21–24, 39, 77, 95, 108]. Thereby, the assumtions made about these two traffic variables undoubtedly impact the obtained channel holding time distribution. Recall, for instance, that Hong and Rappaport [50] conclude that the channel occupancy parameter is *exponential* based on the *initial assumption* that the *call occupancy time is memoryless.* Wang and Fan [108], who as Hong and Rappaport mathematically investigate the channel holding time distribution, conclude that it is *non-exponential.* Recall, however, that in contrast to Hong and Rappaport, Wang and Fan *initial hypotheses for the unencumbered call holding and cell residence time* are that these random variables have *general distributions.* Furthermore, consider again the simulation results obtained by Khan and Zeglache [58]. As noted earlier, the authors explore the channel holding time distribution taking as a starting point for their investigations exponentially distributed cell residence and call holding time random variables. Contrary to Khan and Zeghlache, Spedalieri *et. al* [96] feed their simulations with log-normally distributed call duration. The resultant channel holding time random variable is fitted with hyper-Erlang-$jk$. Thereby, it is important to bear in mind that the conclusions drawn in the overviewed studies depend on the initial conditions. Moreover, we remind the reader and especially practitioners what Yavuz and Leung [112] state about the empirical approaches when assessing the applicability range and accuracy of the studies that try to determine the probability distribution function of different teletraffic variables: "The results are environment dependent, but no assumptions that can be influential are made, as opposed to previous analytical and simulation studies which results are highly dependent on the assumptions made by the authors." Indeed, the strongest aspect of the results obtained by empirical experiments is that although limited to the specific field (cell sites), they are not biased by (sometimes simplifying) hypotheses but rather reflect real, existing conditions.

Considering the overviewed studies it is worth stressing that most of the research works that accept the exponential distribution as a valid one for modelling teletraffic variables indicate the convinience and tractability it offers to the analysis rather then its accuracy in representing the found probability distribution. In the light of this observation, it is important to add that in the relevant literature authors associate the non-exponential teltraffic variables with realistic situations and the exponential one with analytical simplicity.

## 2.4 Conclusions

This chapter introduced the basic system model of land mobile cellular networks. The main performance metrics for system evaluation were specified and a particular emphasis was made on random variables significant for teletraffic engineering.

The overviewed analytical, simulation, and empirical studies yield the conclusion that the classical assumption for exponentially distributed teletraffic variables might not be, in general, the most adequate one. The research on statistical characterisation of pertinent system variables also shows that the memoryless distribution is well suited for discovering general trends and therefore, more appropriate for, as well as more often found in, mathematical modelling. When the classical memoryless assumption is relaxed, more accurate data can be obtained and therefore, it must be used for design, management, and optimisation of mobile cellular networks.

In Part I of the thesis we focus on more *realistic*[15] scenarios. Note that we do not exclude the existance of actual teletraffic conditions in which the exponential distribution is a good approximation but for convenience adopt the term *realistic* in its usual for the teletraffic analysis meaning. Scenarios with memoryless teletraffic variables are widely studied, whereas the non-exponential present more challenging and as explained in Chapter 3 research questions not investigated before.

In Part II of this thesis we adopt the classical hypothesis for exponentially distributed random variables because the goal of the research reported in there is to get insights into system behaviour of next generation broadban mobile wireless networks. To this end, we apply analytical tools as these allow to isolate the phenomenon of interest from the rest phenomena and thus, to clearly determine its impact. For such purposes, as demonstrated by the overviewed studies, the memoryless hypothesis is very suitable.

[15] *Realistic* is broadly used in the teletraffic literature to denote conditions for which the memoryless assumptions do not hold.

# Part I

# Conventional mobile cellular networks

# *3* Admission control based on statistical teletraffic profile estimates

## 3.1 Introduction

The majority of the admission control proposals published in the open literature consider system model with exponentially distributed call duration, cell dwell, and channel holding time as well as Poisson aggregate (new and handoff) arrival process because in general such a model lends itself to a straightful mathematical study. In effect, the negative exponential probability distribution has played a central role as a classical assumption in the mathematical analysis of mobile cellular networks primarily due to its important memoryless property. In the previous Chapter 2 we summarised the main teletraffic variables relevant to the system-level analysis of mobile cellular networks. We centred upon the statistical nature of these random variables by overviewing the works that probabilistically describe them. It was shown through the full arsenal of engineering tools—analytical, empirical and simulation—that in general the teletraffic parameters that characterise a mobile system might not be memoryless by nature and that they are approximated through the negative exponential pattern mainly because of convenience and analytical tractability. In Part I of the thesis we focus on mobile cellular network conditions for which classical assumptions do not hold. In particular, in Section 3.1.2 we overview the works that contribute to the teletraffic analysis of mobile systems for more *realistic* scenarios (i.e., when the exponential hypothesis is relaxed). Then, in Section 3.1.3, we explain the motivation behind our study and specify our contributions in the area. Finally, the last part, Section 3.1.4, outlines the remainder of the chapter.

### 3.1.1 System model

It is essential before proceeding further to recall that the proposals summarised in [5, 31, 42, 57, 88, 104] and those overviewed below consider a traditional mobile cellular system that follows the model described in Section 2.1, Chapter 2, that is: cellular architecture with overlapping coverage areas, handoff mechanism, and intermittently communicating mobile users. In addition, the system is characterised with base station and call resource allocation as follows:

- The BSs are assigned radio resources in a fixed (*fixed channel allocation*, FCA), dynamic (*dynamic channel allocation*, DCA) or hybrid (*hybrid channel allocation*, HCA) manner (see [25, 42, 57, 104] and references therein). The DCA are more efficient compared to FCA schemes in terms of radio resource utilisation but considerably more complex and therefore of less practical interest. The hybrid schemes combine the advantages of both fixed and dynamic schemes. In practice, FCA is often implemented due to its simplicity.
- Each call is assigned a *fixed* amount of resources for the duration of the connection. Moreover, as one service is considered (*voice* or *stremaing*) all accepted call requests are assigned the same amount of fixed resources denominated a *channel*.

That is, the prevailing conditions adopted in the admission control proposals summarised in the aforementioned compendia and considered in Part

I are for deterministic call resource allocation and BS fixed channel allocation (i.e., FCA). Furthermore, it is commonly assumed that the network is homogeneous; that is, consists of identical in terms of size, capacity, and traffic conditions cells. As a result, the system can be studied by modelling the performance of an *isolated cell*, which is the approach adopted in this thesis as well. The impact of interference, fadding, and mechanisms such as power control are omitted, which facilitates the development of pure performance models. The system and the supported voice service are specifed by the parameters as follows:

- $C$ – total number of system resources (channels)
- $\lambda$ – mean of the aggregate call arrival rate given by:

  $$\lambda = \lambda_n + \lambda_h \tag{3.1}$$

- $\lambda_n$ – mean of the new call arrival rate
- $\lambda_h$ – mean of the handoff call arrival rate
- $\alpha$ – mobility factor that shows the average number of handoffs that a call goes through assuming there are infinite resources (no handoff dropping)

  The mobility factor can be calculated considering the mean of the new and handoff arrival rates, in which case it is given by [11]:

  $$\alpha = \left\lfloor \frac{\lambda_h}{\lambda_n} \right\rfloor \tag{3.2}$$

  Note that in analytical modelling the mobility factor is usually derived from the mean call duration and mean cell dwell time (i.e., $\eta/\mu$).

- $1/\mu$ – mean of the unencumbered call duration
- $1/\mu_r$ – mean of the channel holding time

  The mean of the CHT parameter can be determined through the mean duration of the call ($1/\mu$) and the average number of handoffs ($\alpha$), in which case it is given by:

$$1/\mu_r = \frac{1/\mu}{\alpha + 1} \tag{3.3}$$

- $A$ – offered traffic load given by:

$$A = \frac{\lambda}{\mu_r} \tag{3.4}$$

### 3.1.2 Literature that relaxes the classical teletraffic assumptions

The [5,31,42,57,88,104] compendia provide a summary of the vast research carried out in the area of call admission control in mobile cellular networks. Here, we overview the proposals from the perspective of the goals defined below. Specifically, we focus on the literature that relaxes the classical teletraffic hypotheses for exponentially distributed random variables.

Motivated by the statistical results concerning the probabilistic behaviour of teletraffic random variables of interest (see Chapter 2) some authors investigate system performance for realistic assumptions and (or) develop anaytically tractable models that can be used in the design and evaluation stages of mobile cellular networks. Rajaratnam and Takawira [78] for example, propose a mathematical framework for studying the performance of the *guard channel scheme*[1] under smooth handoff traffic and gamma and det-neg channel holding time distributions. The authors indicate that there are not any noticeable differences between the new call blocking probabilities for different channel holding time cases ("no dependence beyond the

[1] The *guard channel* (or *cut-off*, *trunk reservation*) scheme is a cornerstone in mobile cellular admission control research. The cut-off concept has not only been widely studied by the research community for a variety of traffic conditions and assumptions but also extensively used in the design of admission control algorithms. The guard channel scheme divides system capacity into a *common* and *reserved* (*guard*) pool of channels. The former is accessible to all call classes (new and handoff), whereas the latter – only to the handoff calls. Active calls are given higher priority (as common to all AC schemes): handoff requests are always accepted provided there are free channels in the system. New call requests are admitted only if there are enough free resources in the common pool of channels. The common and guard pools refer to amount of channels.

mean", [78]). The reason for this is that the new call arrival traffic flow is modelled as a Poisson process and Poisson arrivals are insensitive to service time distributions beyond the mean. Rajaratnam and Takawira, however, stress that the handoff dropping probability does depend on the service time distribution when smooth handoff arrival traffic is considered and elaborate on the relation betweent the squared coefficient of variations ($SCV$) of the channel holding time and the dropping probability: the lower the $SCV$ of the service (channel occupancy) time, the lower is the dropping probability [78] and vice versa. Dharmaraja *et al.* [30] analyse system performance under exponentially distributed channel and cell holding time parameters, Poisson new and general handoff traffic. Dharmaraja *et al.* [30], as Rajaratnam and Takawira [78], focus on the guard channel scheme and demonstrate that the forced termination probability depends on the handoff arrival distribution. For Erlang handoff arrival process the authors show that the dropping probability is decreased in comparison with the one registered when the handoff traffic is Poissonion but that the dropping probability is considerably degraded (i.e., increased) for hyper-exponentially distributed handoff interarrival times ($SCV > 1$). The results also demonstrate that when the total number of channels (BS capacity) is augmented, the dropping probabilities for the case when Poisson and the case when non-Poisson handoff traffic flows are modelled tend to approach each other. The latter is due to the fact that when system capacity is significantly rised (more resources for serving arriving requests) the conditions approach a non-blocking environment; therefore, the traffic is not distort and keeps its Poisson distribution. The results obtained by Dharmaraja *et al.* [30] show that the new call blocking probability is not altered significantly by the handoff distribution parameters; that is, the same blocking probability is reported for Poisson and non-Poisson handoff arrival processes.

Khan and Zeghlache [58] and Chlamtac *et. al* [18] evaluate mobile cellular system performance for exponential call holding time, general cell residence time, and non-Poisson handoff arrival process. It is shown that in such conditions the PASTA[2] property is not applicable and can not be used for evaluating the handoff traffic performance. Moreover, it is de-

[2] PASTA stands for *Poisson arrivals see time averages.* According to the PASTA property the time average is equal to the event average.

monstrated that handoff call dropping depends on the variance of the cell residence time and user mobility while the new call blocking probability is only slightly altered; that is, the new and handoff call arrival traffic experience different blocking. According to Chlamtac *et. al* [18], their results are a much closer approximation to measured field data than the studies that assume exponential distribution or make simplifying assumptions. Likewise, Khan and Zeghlache [58] show that there exist prominent dependancies of handoff performance metrics on cell residence time distribution and that the exponential one can be mainly used in analytical studies for determing general trends. Furthermore, Zeng and Chlamtac [113] among several others emphasise that handoff system performance depends on the channel holding time through the cell residence time distribution. Orlik and Rappaport [71] model system performance under more general conditions as well. Specifically, the call holding and cell dwell time distributions are modelled through the proposed by the same authors sum of hyper-exponentials distribution. For these conditions, the effect of the variance of the distribution (of both holding and residence times) is manifested in the handoff rate and dropping and is less clearly seen in the new call blocking. Pattaramalai *et al.* [32, 72] derive the call completion probability for Weibull [72] and general [32] cell residence and channel holding time distributions. The authors affirm that the call completion metric depends largely on the shape parameter of the aforementioned cell dwell time and channel holding time distributions [32, 72].

A fundamental result common to the overviewed works is that system performance depends on the teletraffic profile. Motivated by this central conclusion researchers such as Fang *et. al* [37], Orlik and Rappaport [70,71], Corral-Ruiz *et. al* [22–24], intend to propose a general, unified and feasible analytical framework that allows performance evaluation of mobile cellular networks for non-trivial assumptions.

Note that the overviewed publications focus on the *effect* that the non-exponentially distributed teletraffic variables have on system-level performance. They demonstrate that for the accurate design, operation, and management of mobile cellular networks the statistical characterisation of these parameters is essential. However, to the best of the knowledge research that investigates the applicability of the statistical properties of these variables to devising admission control schemes has not been carried out (or at least not reported) in the open literature. The research work, of which we give a written account in this chapter as well as in the remainder of Part I, intends to fill this gap.

### 3.1.3 Contributions

The research work reported in Chapter 3 was strongly motivated by the fact that despite extensive investigations into the probability characterisation of cell residence, call duration, and channel holding time, the results from these studies had not been considered for the development of admission control algorithms. Furthermore, we conjectured that it would be convenient to implement the intrinsic statistical properties of the random variables discussed in Chapter 2 into admission control. Our main objective thus, was to answer the interesting and practical question whether the *statistical profile* of the teletraffic variables can be *advantageously exploited* so that *efficient yet simple admission control* algorithms can be designed, assuming that we know their statistical profile. In this context, the main contributions of our research work are as follows:

- We *propose* an admission control algorithm for traditional voice–oriented mobile cellular networks based on the channel holding time *statistical profile.* Compared to the classical cut-off (guard channel) scheme, the proposed algorithm supports a much *wider* (continuous in contrast to discrete) *working interval*, which gives mobile operators the freedom to choose the desired trade-off between QoS and revenue (i.e., blocking probabilities and carried traffic). At the same time, the proposed admission control scheme has practical value because its computation and implementation is *straightforward* as it uses simple statistical metrics that can be readily estimated in the system. As explained earlier, wireless network operators choose admission control schemes on the basis of their simplicity (complexity) of implementation and overall performance.

- We mathematically *analyse* the proposed admission control scheme for *classical memoryless hypotheses.* In particular, we assume that all the observed teletraffic processes follow a n.e.d., which is a general working hypothesis accepted in the majority of analytical works. We show that for such assumptions the cut-off scheme is a special case of the proposed algorithm.

- We *evaluate* the proposed scheme performance for the more cumbersome case of *general channel holding time distribution.* We develop a simulation tool that allows system-level evaluation for such non-trivial traffic conditions and probability distributions. The implemented simulation tool was used in the subsequent research (see Chapters 4 and 5, Part I).

- The *overall contribution* of the research carried out in Part I of this thesis is that we show for the first time that the intrinsic statistical properties of the teletraffic variables pertinent to mobile cellular networks can be efficiently exploited in the development of admission control algorithms and that such AC schemes bring improvements in system performance.

### 3.1.4 Outline

The remainder of Chapter 3 is structured as follows. In Section 3.2 we briefly overview the fundamental concepts of renewal theory and explain the mathematical reasoning used to explore the statistical properties of the teletraffic variables overviewed in Chapter 2. Next, we elaborate on the particular algorithm that we propose (see Section 3.3). Then, we develop the methodology we follow to examine the proposed scheme (see Section 3.4). In Sections 3.5 and 3.7 the suggested methodology is put into practice. Specifically, in Section 3.5 the proposed admission control is analysed adopting conventional hypotheses. Later, in Section 3.7, after explaining in Section 3.6 the simulation tool that we developed, we evaluate system performance for general probability density function conditions. Lastly, Section 3.8 concludes the chapter by summarizing our main findings.

## 3.2 Background

We overview renewal theory to provide the very basic concepts that are used in the derivation and definition of the residual lifetime parameter. The latter has been intensively used primarily by statisticians in a wide range of applications. In reliability theory (applied in insurance, manufacturing, monitoring, and industry in general) for example, the main concern is the frequency with which fault in the item is observed or for how long a monitored item can survive. Other areas of application of the residual lifetime metric are social science, different branches of the medicine (e.g., epidemiology), biostatistics, etc., and in general all areas of human activity where estimation of residual time of a random variable is relevant. Closest to our study is the application of the residual lifetime in queueing systems, where the expected residual lifetime is used to estimate the time that an arriving customer has to wait until the busy service facility is freed (see Chapter 5, Section 5.2 in [61]).

Recall that a renewal process is a counting process $\{N(t),\, t \geq 0\}$ for which the times $X_k$ between successive events $(E_i,\, E_{i+1})$ are independent and identically distributed with an arbitrary probability distribution [83]. Let the cumulative distribution function of the independent and identically distributed time intervals $X_k = t_k - t_{k-1}$ is given by:

$$F(t) = P[t_k - t_{k-1} \leq t]$$

and their probability density function is given by:

$$f(t) = \frac{dF(t)}{dt}$$

Let also $m_1 = E[X_n]$, $n \geq 1$ denote the mean of the time intervals between successive events.

Consider now the point in time $\tau$, randomly chosen, and the time interval (that we call *sample* interval as in [61, 83]) within which $\tau$ falls; these are sketched in Fig. 3.1. In renewal theory the lenght of the time interval $X_n$ is denoted *lifetime*, $X = \tau - t_n$ is called *age*, and $Y = t_{n+1} - \tau$ is the *excess* (or *residual*, *remaining*) *lifetime* at time instant $\tau$. The probability

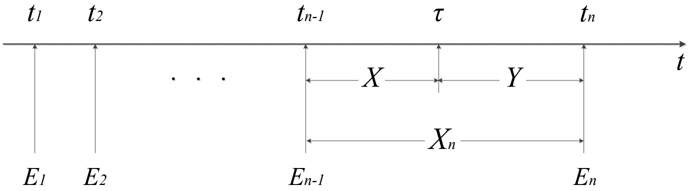


Figure 3.1: Renewal process

density function of the sampled interval $X_n$ is given by $tf(t)/m_1$, that is, in terms of the density of typical intervals and their mean.

The renewal theory provides us with several useful results about the excess lifetime. In particular, assuming that the ponit of time $\tau$ is uniformly distributed within the sampled interval (recall that it was randomly chosen from the time axis), the probability density function of the residual lifetime is given by [61, 83]:

$$\hat{f}(t) = \frac{1 - F(t)}{m_1}, \tag{3.5}$$

where

$$F(t) = \int_0^t f(y)dy,$$

and $m_1$ is the first moment of the interarrival lenghts as specified previously.

Another significant statistic is the mean of the residual life, which is given by [61, 83]:

$$r_1 = \frac{E[X^2]}{2E[X]} = \frac{m_2}{2m_1} = \frac{m_1}{2} + \frac{\sigma^2}{2m_1}$$

An extensively used in the practice function is the age-dependent *failure rate* $r(t)$ defined as $r(t)dt = P\{t < \text{lifetime} \leq t + dt \,|\text{lifetime} > t\}$; that is, the instantaneous rate at which an item will fail given that it has already attained age $t$ [83]:

$$\begin{aligned} P\{t < X < t + dt|X > t\} &= \frac{P\{t < X < t + dt,\, X > t\}}{P\{X > t\}} \\ &= \frac{P\{t < X < t + dt\}}{P\{X > t\}} \\ &\approx \frac{f(t)dt}{1 - F(t)} = r(t)dt \end{aligned}$$

The conditional probability density function that a $t$-year-old item will fail then is given by:

$$r(t) = \frac{f(t)}{1 - F(t)} \tag{3.6}$$

where $f(t)$ and $F(t)$ are as defined previously (i.e., the common probability density and cumulative distribution functions of the interarrival time intervals respectively).

Another statistical metric of interest is the density of the excess lifetime of a $x$-year-old item. The probability density function of the remaining time (excess lifetime $Y$) provided that age $X = \epsilon$ has been attained is given by [10]:

$$\hat{f}(t|\epsilon) = \frac{f(t + \epsilon)}{1 - F(\epsilon)} \tag{3.7}$$

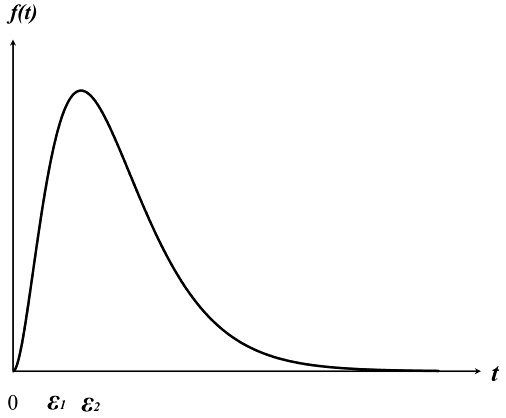
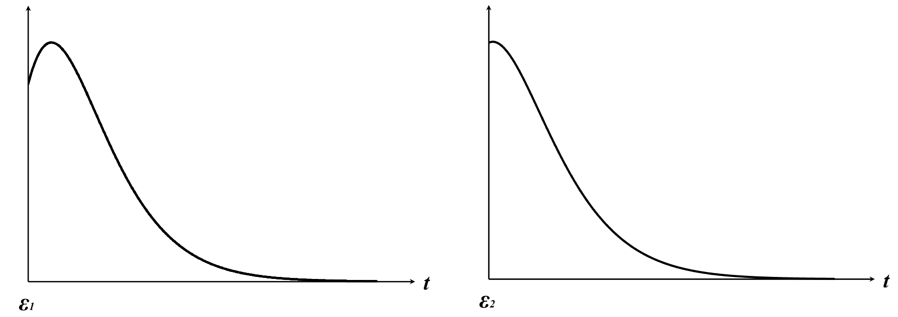


Figure 3.2: Probability density function of a random variable with a squared coefficient of variation smaller than one

To motivate the derivation of the *conditional probability density function of the residual lifetime*, eq. (3.7), let us overview the reasoning of Barceló [10]. Since the interest is in the case when the item has reached age $X$ and this age is known, i.e., $X = \epsilon$, the probability density function of the lifetime from that very instant on should be taken into account; that is, the density function of lifetime intervals longer than $\epsilon$. Consider, for instance, the probability density function[3] shown in Fig. 3.2 of the interarrival lengths sketched in Fig. 3.1. Assume now that some time has elapsed from the beginning of the time interval and that the actual age $X$ is equal to $\epsilon_1$. The probability density function of the residual lifetime then is determined by the curve shown in Fig. 3.3 (a). That is, the conditional density of the residual lifetime is obtained from the probability density function of the interarrival times from the instant $\epsilon_1$ on, excluding lifetime intervals $X_i$ shorter than $\epsilon_1$. Likelise, when the elapsed time is $\epsilon_2$, the probability

[3] Note that the same reasoning is valid for the case of a probability density function with $SCV > 1$. A probability density function with $SCV < 1$ is plotted because it is easier to appreciate the sketch of its residual lifetime density function.



**(a)** $X = \epsilon_1$ **(b)** $X = \epsilon_2$

Figure 3.3: Conditional probability density function of the residual lifetime of the random variable sketched in Fig. 3.2 when the attained age is $X = \epsilon_i$, where $\epsilon_1 < \epsilon_2$

density function of the remaining time $\hat{f}_2(t|\epsilon)$ is determined by intervals longer than $\epsilon_2$ and is sketched in Fig. 3.3 (b). Furthermore, recall that a probability density function $f(x)$ by definition must satisfy the condition:

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx$$

In effect, the probability density function $f(t)$ of the lifetime intervals (intearrival lengths) $X_i$ satisfies the condition:

$$1 = P\{T \in [0, \infty)\} = \int_{0}^{\infty} f(t)dt$$

Thereby, in order to meet the requirement for the area under the probability density curve to be unity, that is:

$$\int_{\epsilon}^{\infty} \hat{f}(t|\epsilon)dt = 1,$$

the lifetime density function $f(t+\epsilon)$ is normalized by $1 - F(\epsilon)$, where

$$F(\epsilon) = \int_{0}^{\epsilon} f(t)dt$$

Note that there is a subtle difference between the discussed probability density functions (see equations (3.5) and (3.7)). The density function of the residual lifetime when the elapsed time is *uniformely distributed* is given by equation (3.5), whereas equation (3.7) concerns the density function of the residual lifetime when the attained age $X$, i.e., the elapsed time since the beginning of the interval, is *known*; that is why we call the latter **conditional probability density fucntion of the residual lifetime**. Recall as well one of the main results of renewal theory, namely the instantaneous failure rate, $r(\epsilon) = f(\epsilon)/1 - F(\epsilon)$. The latter is the conditional density function of the *instantaneous rate* at which an item fails, whereas eq. (3.7) gives the conditional denisty function of the *residual lifetime*.

The expected value of the residual lifetime provided that $X = \epsilon$ age was attained, which we denote $\bar{h}(\epsilon)$, is defined through the probability density function $\hat{f}(t|\epsilon)$ and is given by:

$$\bar{h}(\epsilon) = \int_{0}^{\infty} t\hat{f}(t|\epsilon)dt \tag{3.8}$$

In the next section we elaborate on how this statistical estimate can be used as a parameter in admission control.

## 3.3 Admission control based on statistical teletraffic profile estimates

### 3.3.1 Concept

The main idea that we propose is based on two pillars: the scientific evidence about the statistical nature of the teletraffic processes observed in mobile cellular networks *and* the main results of renewal theory and in particualar, the fundamental concept exploited in reliability theory and practice: decision making based on the estimated system (item, server, client, etc.) behaviour. Another important factor that we considered is that the first two moments of the arrival process, call holding, and channel holding time are easily obtained in the BS; hence, their implementation in admission control was expected to be easy and feasible. We took as a starting point the case when the distributions of these random variables are known. Later, we relaxed this assumption as well.

Let us go back to the central conclusion derived in Chapter 2. As explained there, the statistical results suggest that in general the exponential distribution is not the best fit to empirical data, which conclusion is supported by analytical and simulation studies that relax conventional hypotheses. As a consequence, the future system state cannot be estimated based only on the present state of the system but depends on the past evolution of the system. The latter has been considered and indeed is a major impediment in analytical modelling as it significantly complicates the analysis and evaluation of mobile cellular systems. However, we examine the non-memoryless properties of the teletraffic variables from a perspective that allows us to benefit from these probabilistic properties.

Specifically, we denote with *remaining time* the time interval between the instant when a decision for a call request admission (rejection) has to be made and the next renewal point. Renewal can be the instant when a new call arrives or a resource is released. As noted previously, if the density function of the call interarrival or channel occupancy times is known, then the mean residual lifetime (that is, the mean remaining time until a consecutive request arrives or a channel becomes free due to call termination

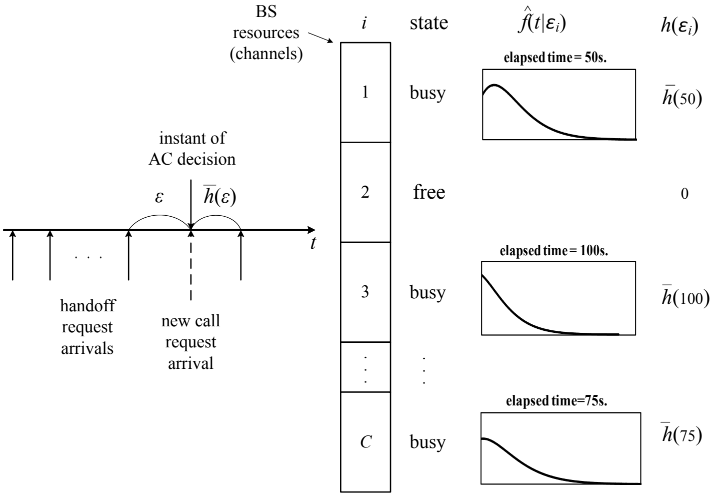


Figure 3.4: System state upon new call arrival, that is, at the instant when admission control (AC) decision must be made: mean remaining handoff interarrival $h(\epsilon)$ and channel occupancy times $\bar{h}(\epsilon_i)$ at that very instant

or continuation of the call in a neighbouring cell) can be easily obtained. These estimates can be used as a (powerful) tool in devising admission control algorithms. For example, the decision of accepting a new call request can be based on the current system state (number of free channels, estimated channel release, etc.) and the expected remaining time until a handoff (that is, high priority) call arrives (see Fig. 3.4). Several admission control schemes that implement similar metrics can be divised, and later on we elaborate on a particular one. It should be noted here that in contrast to the schemes with fixed resource reservation such as the guard channel one, admission control algorithms based on the aforementioned parameters are *flexible* by nature: the decision is based on present and estimated future system state rather than on the current resource occupancy only. The latter is considered advantageous to system performance as system resources are not reserved in a fixed manner but dynamically, depending on current and estimated future conditions.

### 3.3.2 MRT Algorithm

We first begin with an intuitive explanation of the algorithm that we developed and than proceed with a more rigourous and detailed specification.

The admission control algorithm that we propose is based on the measured statistical behaviour of teletraffic random variables as noted in the preceding section. This is in contrast to the common approach in which exponential *assumptions* are made. We suggest to use a metric that is related to the channel holding time variable. We are interested in determing for how long the occupied channel will remain busy; that is, in the mean remaining time of an ongoing call in the cell (BS) given that the call has already been in progress (assigned a channel by the base station) during a certain time. Note that the elapsed time from the beginning of the service is known at the BS – such data is readily available as it is needed for billing purposes for example. Based on the time elapsed from seizing a free channel, the remaining time for releasing it can be probabilistically estimated. The mean remaining time until a channel is freed in the system can be used to decide on the acceptance of a new traffic. If, for instance, the estimated residual occupancy time is long, it can be concluded that the system will remain busy and therefore the new call shall be rejected in order to avoid future system congestion. If, on the contrary, based on the estimate, it is expected that a channel will be freed soon the new call request can be admitted into the system as its acceptance will not lead to system overload. Avoiding system congestion is crucial otherwise, the system cannot allocate resources to arriving ongoing call requests and as a result, these calls are interrupted.

The proposed admission control scheme accepts the arriving handoff call requests, provided that there are available resources. A random variable, which estimates the expected time until a busy channel is released, determines whether new traffic is admitted (see Fig. 3.5). Once a call of any type (that is, new or handoff) has been accepted, it can use *any* free channel. Below we provide a detailed specification of the algorithm.

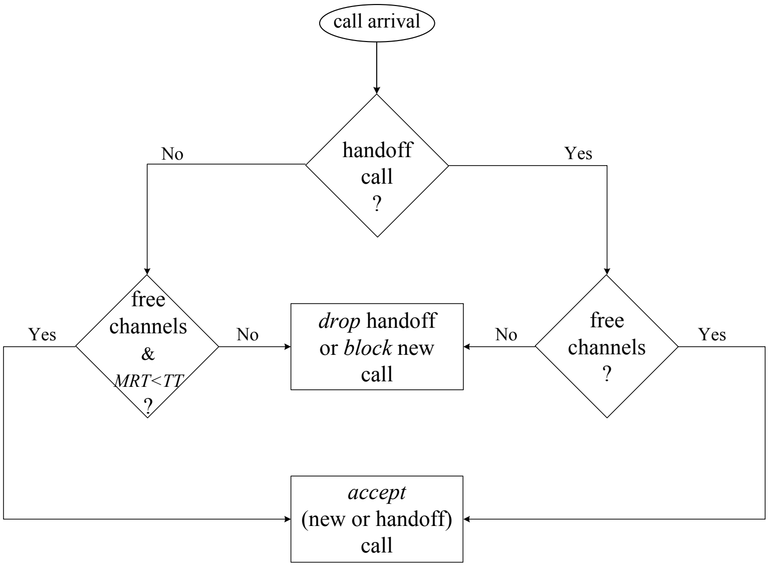


Figure 3.5: Logic of the proposed teletraffic-based handoff method

### Handoff Calls

A handoff call must be served whenever possible because users perceive a forced call termination as more inconvenient than a blocked call initiation. The acceptance probabilities of handoff calls are therefore given by:

$$p_{ho} = \begin{cases} 1 & \text{if free channel} \\ 0 & \text{if system busy} \end{cases}$$

This procedure of always accepting a handoff call request, provided there are enough free resources (see Fig. 3.5), is common to all admission control algorithms. The reason for this is that by giving higher priority to handoff requests call interruptions are minimised[4].

[4] The admission control algorithms discussed in the thesis are often called in the literature *handoff priority* schemes as ongoing calls are given priority over new call requests.

### New Calls

New call requests are accepted if two conditions are simultaneously satisfied: (1) there is a free channel and (2) admission control criterion is met. This way, pre-existing (*active, ongoing*) call connections are prioritised over new call requests. The main idea, as explained previously, is to block new traffic before congestion occurs, in order to guarantee that there will be free resources in the base station upon arrival of a handoff call. The acceptance of new traffic depends on a random variable, which probabilistically estimates the expected time until the next release of a busy channel. We define this random variable, which is herein denoted *mean remaining time* ($MRT$), to be given by:

$$MRT = \frac{1}{C}\sum_{i=1}^{C}\overline{h}(\epsilon_i) \tag{3.9}$$

where $C$ is the total number of channels (system resources) and $\overline{h}(\epsilon_i)$ stands for the mean remaining channel occupancy time ($\overline{h}$) of an ongoing call that is using channel $i$, given that the elapsed time since the call has occupied the resource $i$ is $\epsilon_i$. The mean remaning channel occupancy time of each individual call $\overline{h}(\epsilon_i)$ is given by:

$$\overline{h}(\epsilon_i) = \begin{cases} \int_0^{\infty} t\hat{f}(t|\epsilon_i)dt & \text{if} \quad j_i = 1 \\ 0 & \text{if} \quad j_i = 0 \end{cases} \tag{3.10}$$

where $\hat{f}(t|\epsilon_i)$ is, as noted earlier, the conditional probability density function of the remaining resource holding time given that channel $i$ was occupied $\epsilon_i$ time units before and the state $j$ of channel $i$ is denoted as follows:

$$j_i = \begin{cases} 1 & \text{if channel } i \text{ is busy} \\ 0 & \text{if channel } i \text{ is free} \end{cases}$$

Note that when the channel is free we set the mean remaining time $\overline{h}(\epsilon_i)$ to zero but include it into the calculation of the $MRT$ estimate (3.9) to account for currently available resources.

Furthermore, the acceptance of new calls is regulated through a *time threshold* ($TT$). The estimated mean remaining time until a channel is released must be less than the $TT$ for the new call to be accepted. The acceptance probabilities of new calls then are given by:

$$p_{new} = \begin{cases} 1 & \text{if} \quad MRT < TT \\ 0 & \text{if} \quad MRT \geq TT \end{cases}$$

The time threshold $TT$ is used to block new calls with anticipation, in order to guarantee the successful continuation of arriving handoff call requests. If the $TT$ is restrictive, then a new call will only be admitted given that a channel is expected to be freed within a short time interval. This admission restriction probabilistically ensures that upon a handoff request arrival, the ongoing call will not be interrupted due to lack of a free channel. Hence, the probability that the BS will successfully serve ongoing calls will increase. In contrast, a longer $TT$ will make the BS more receptive to incoming new call requests but will decrease the probability of channel availability: that is, it will reduce the probability of admitting and thereby, successfully carrying out, handoff call requests.

Note that $TT \in [0, \infty)$. Moreover, when $TT = 0$ only handoff calls are accepted, whereas when $TT \to \infty$ (in practice long enough) the system accepts all incoming call requests, that is, there is no prioritisation.

The proposed algorithm, hereafter denoted MRT scheme for short, implements the intrinsic charasteristics of the mobile cellular CHT random variable. At the same time the used statistical metrics of the channel holding time are simple and easy to estimate on-line[5]. In particular, the mean and variation of the channel holding time can be estimated along time windows that are long enough to allow averaging and short enough to assume that the traffic processes are stationary. When the normal duration of connections and ITU-T recommendations on teletraffic are considered, this window is typically one hour. Furthermore, the MRT algorithm is straightforward to compute, which is essential for practical implementation. As

[5] Recall that we assumed that the channel holding time distribution is known. Our conjecture that operating knowing only the first two moments, without a precise knowledge of the distribution, is studied later in Chapter 5.

stressed earlier, the evaluation of the admission control criteria for accepting (rejecting) a new call request is executed *frequently* (in practice, upon call request arrival; depends on the offered traffic load but could be in the order of mili- or micro-seconds) and must be computed *quickly* (in the order of micro-seconds as well). Another advantage of the algorithm is that it does not load the system with additional control traffic and is of local scope; that is, it does not require interchange of control data between BSs.

## 3.4 Methodology

### 3.4.1 Approach

The methodology that we established for systematically studying the MRT scheme is as follows. First, we examined the proposed algorithm for conventional working hypotheses, namely exponentially distributed random variables and Poisson arrival traffic. The resulting system model was trivial as product-form solutions were readily available (see Section 3.5). Then, we relaxed the hypothesis regarding the channel holding time distribution but retained the remainder of working assumptions (see Section 3.7). The latter, allowed for isolating the effect of the channel occupancy parameter from the rest of the variables and clearly assessing its impact on system performance. Later, the same hypotheses were used but system performance with MRT scheme was evaluated for more realistic environment (see Chapter 4, Section 4.3). This included handoff areas and high mobility, which are conditions typical for present and are expected to be common to future broadband mobile cellular networks. The next step in studying the MRT scheme was to relax the Poisson arrival assumption, which is done in Chapter 4, Section 4.4. Finally, we relaxed the hypothesis for knowing the channel holding time distribution and examined the MRT scheme performance for such a restrictive condition (see Chapter 5).

We considered the classical guard channel scheme as a reference case and compared MRT scheme performance to it, which is a common approach in admission control research (see [5,31,42,57,88,104] compendia)[6].

[6] Note that another broadly used approach is to compare system performance when a

### 3.4.2 CHT case studies considerations

In order to investigate the MRT scheme for the complete range of possible channel holding time conditions we differentiate between channel holding time scenarios based on the value of the *squared coefficient of variation* ($SCV$) of the CHT distribution. Recall that the $SCV$ is given by [9]:

$$SCV = \frac{\sigma^2}{m_1^2}$$

where $\sigma$ is the standard deviation of the random variable and $m_1$ is its mean. Note as well that the exponential distribution is the only one for which $SCV = 1$. Thereby, the squared coefficient of variation is often used to measure the irregularity of a random variable in comparison with the exponential one [9]. When the hypothesis for exponentially distributed channel holding time variable is relaxed we can distinguish two general cases based on the $SCV$: *hypo-exponential* ($SCV < 1$) and *hyper-exponential* ($SCV > 1$). We based our case studies on the $SCV$ as a main criterion because it is not feasible to measure system performance for all the possible channel holding time distributions, first and second moments. Moreover, we argue that our approach embraces cases general enough to capture system performance for the complete range of statistical conditions.

Specifically, we focused first on exponentially distributed CHT and then on the non-memoryless CHT. We chose *Erlang*-3, which has a $SCV = 1/3$, and *balanced hyper-exponential* (herein denoted HE2b) with $SCV > 1$, to exemplify the principal idea and functionality of the MRT scheme when the $SCV \neq 1$. These two distributions exhibit contrasting behaviour, which enables a broader spectrum of conditions to be tested. In particular, the average remaining time for the Erlang-3 distribution is a monotonically decreasing function: the longer the elapsed time of a call in a base station,

proposed admission control algorithm is used with system performance when no admission control is implemented. Such evaluation study gives a notion of the absolute improvement introduced by the proposed scheme into the system. A comparison of the MRT scheme with the cut-off concept shows the relative improvement due to the proposed scheme. Naturally, the absolute improvement is larger that the relative one.

the higher the probability that the channel will soon be released. Conversely, the mean remaining time of HE2b distributed channel holding time is a monotonically increasing function. That is, the longer the elapsed time of an ongoing call in the system (BS), the longer the expected time it will remain in service (occupying the allocated resource). To elaborate further, let us focus on the case when the channels in a base station have been seized a long time ago. If the channel holding time is Erlang-3 distributed, then the longer the elapsed time in the busy channels, the higher the probability that a channel will be released and therefore the higher the probability that a new call request will be accepted. As a result and due to the logic of the MRT scheme, more new calls will be accepted (as the system is expected to have a lot of free channels shortly after). If the channel hodling time is HE2b distributed however, the greater the number of channels seized long time ago, the greater the mean remaining time ($MRT$) until a channel will be freed. Consequently, more new calls will be blocked by the system according to the MRT scheme's logic. Hence, we assert that by investigating the MRT scheme under the chosen conditions, a comprehensive performance evaluation can be carried out.

Another fact that we took into consideration while elaborating on the particular case studies was that the aforementioned distributions have been used in modelling mobile systems (see Chapter 2). In practice, the Erlang-$k$ distribution was used to fit real teletraffic variables in cellular systems and in analytical studies[7] as described in Chapter 2. The phenomenon of two call classes – calls that are originated and remain in the same cell (*new*) and calls that move to a neighbouring cell (*handoff*) – observed in mobile cellular networks, has been modelled by probabilistically combining two exponential random variables with different means, which in effect is a hyper-exponential random variable [54]. Moreover, these distributions can be decomposed into memoryless stages (see below), which memoryless property could be used in an analytical study of the algorithm.

[7] F. Yang uses Erlang-$k$, whereas Rajaratnam and Takawira use gamma distribution for modelling certain teletraffic parameters. Recall that the Erlang distribution is a special case of the gamma distribution when the shape parameter of the latter is an integer.

### 3.4.3 CHT case studies

We summarise below the main metrics relevant to the (remaining) channel holding time random variable for the three studied $SCV$ cases.

#### Case study I: exponentially distributed CHT

Recall that a random variable $T$ is said to be without memory (or *memoryless*) if [83]

$$\begin{aligned} P\{T > x + t | T > t\} &= \frac{P\{T > x + t, T > t\}}{P\{T > t\}} \\ &= P\{T > x\} \, \forall \, x, \, t \geq 0 \end{aligned}$$

which is equivalent to

$$P\{T > x + t\} = P\{T > x\} P\{T > t\}$$

If $T$ is the channel holding time (or the lifetime of the interarrival times as in Section 3.2), then the preceeding equation states that the probability that the call occupies the channel for at least $x + t$ time units given that it has been in service for $t$ time units is the same as the initial probability that the call will be served for at least $x$ time units. In other words, if the channel is busy at time $t$, then the distribution of the remaining amount of time that the call will occupy the channel is the same as the original channel holding time distribution [83]. That is,

$$\hat{f}(t|\epsilon) = f(t) \tag{3.11}$$

where the probability density function of the continuous random variable $T$, which has an exponential parameter $\mu > 0$, is given by:

$$f(t) = \begin{cases} \mu e^{-\mu t} & t > 0 \\ 0 & t \leq 0 \end{cases}$$

The exponential variable $T$ with parameter $\mu$ has mean $m_1 = 1/\mu$.

### Case study II: Erlang-3 distributed CHT

Recall that Erlang-$k$ random variables are a special class of gamma r.v. named after the Danish mathematician and teletraffic engineer A. K. Erlang. A random variable $T$ has Erlang distribution with parameters $k$ and $\mu$ and density function given by:

$$f(t) = \begin{cases} \frac{\mu k(\mu kt)^{k-1}}{(k-1)!} e^{-\mu kt}) & \text{for} \quad t > 0 \\ 0 & \text{for} \quad t \leq 0 \end{cases}$$

The physical meaning of an Erlang-$k$ random variable is of passing through $k$ identical and independent stages, each of which with an exponential time with parameter $\mu k$.

The probability density function $f(t)$ of Erlang-3 random variable with mean $3/\mu$ is given by:

$$f(t) = \frac{1}{2}(\mu^3 t^2 e^{-\mu t})$$

The conditional probability density function of the residual lifetime of Erlang-3 random variable when the elapsed time is known is derived applying (3.7) and is given by:

$$\hat{f}(t|\epsilon) \quad = \quad \frac{1}{1 + \mu\epsilon + (\mu\epsilon)^2} \left( \frac{1}{2}(\mu^3 t^2 e^{-\mu t}) + \frac{1}{2}(2\mu^3 t\epsilon + \mu^3 \epsilon^2) e^{-\mu t} \right)$$

Note that the first term of the sum is in practice the probability density function of Erlang-3 random variable. A more compact expression of $\hat{f}(t|\epsilon)$ is given by:

$$\hat{f}(t|\epsilon) = \frac{1}{2} \frac{\mu^3 (t + \epsilon)^2 e^{-\mu t}}{1 + \mu\epsilon + \frac{(\mu\epsilon)^2}{2}} \tag{3.12}$$

The mean of the residual lifetime when the elapsed is known and equal to $\epsilon$ is given by:

$$\overline{h}(\epsilon) = \frac{6 + 4\mu\epsilon + (\mu\epsilon)^2}{2\mu + 2\mu^2\epsilon + \mu^3\epsilon^2} \tag{3.13}$$

### Case study III: HE2b distributed CHT

In contrast to the Erlang-$k$ random variable, which can be seen as a series of $k$ identical exponential stages, the hyper-exponential random variables consists of two parallel stages, each of which is exponentially distributed with parameter $\mu_i$, $i = 1, 2$. The stages are entered with probability $p$ and $1 - p$ respectively. The mean of the hyper-exponential distribution is given by:

$$\frac{1}{\mu} = \frac{p}{\mu_1} + \frac{1-p}{\mu_2}$$

When the mean is balanced, that is:

$$\frac{p}{\mu_1} = \frac{1-p}{\mu_2}$$

the hyper-exponential distribution is called balanced, and we denote it HE2b. The probability density function of a balanced hyper-exponential random variable is given by:

$$f(t) = p\mu_1 e^{-\mu_1 t} + (1-p)\mu_2 e^{-\mu_2 t},\ t \geq 0$$

The conditional probability density function of the residual lifetime when the elapsed time is known is derived from (3.7) and is given by:

$$\hat{f}(t|\epsilon) = \mu(pe^{-\mu_1 t} + (1-p)e^{-\mu_2 t}) \tag{3.14}$$

where $1/\mu$ is the mean of the HE2b proability density function.

The mean of the conditional density function of the residual lifetime when the elapsed time is $\epsilon$ is given by:

$$\overline{h}(\epsilon) = \frac{1}{\mu_1}\frac{pe^{-\mu_1\epsilon}}{pe^{-\mu_1\epsilon} + (1-p)e^{-\mu_2\epsilon}} + \frac{1}{\mu_2}\frac{(1-p)e^{-\mu_2\epsilon}}{pe^{-\mu_1\epsilon} + (1-p)e^{-\mu_2\epsilon}} \tag{3.15}$$

Before proceeding with the analysis and performance evaluation of the MRT scheme recall that its functionality is not limited to a specific scenario. That is, although we evaluate the performance of the algorithm for controlled conditions – channel holding time distributions, which might deviate from those found in practice – these capture the range of all possible CHT cases.

## 3.5 Analysis of the MRT scheme for classical assumptions

In this section we analyse the MRT scheme for conventional working hypotheses, namely we assume that the teletraffic random variables described in Section 3.1 follow negative exponential distributions.

In the preceeding section we noted that the remaining lifetime of a variable with exponential distribution and mean $m_1$ is independent of the already elapsed time and therefore, equals the mean $m_1$. Hence, if we assume that the channel holding time is exponentially distributed then, the estimation of the remaining channel occupancy time does not depend on the elapsed time. Furthermore, the busy channels will have the same mean remaining time $\overline{h}$ independent of the elapsed time $\epsilon_i$ in each busy channel $i$, that is:

$$\overline{h}(\epsilon_i) = \overline{h} = \frac{1}{\mu_r} \tag{3.16}$$

where $1/\mu_r$ is the mean channel holding time (i.e., $m_1 = 1/\mu_r$). Consequently, the mean remaining time $MRT$ will be independent of the instant of time when it is computed but will only depend on the number of busy channels. In effect, when the channel occupancy time is exponentially distributed, $MRT$ is given by:

$$MRT = \frac{1}{C} \cdot \frac{1}{\mu_r} \sum_{i=1}^{C} j_i \tag{3.17}$$

where $j_i$ is as defined earlier; that is,

$$j_i = \begin{cases} 0 & \text{if channel } i \text{ is free} \\ 1 & \text{if channel } i \text{ is busy} \end{cases}$$

or equivalently

$$MRT = \frac{1}{C} \cdot \frac{BC}{\mu_r}$$

where $BC$ denotes the number of busy channels and $C$ is the total number of channels as noted earlier. If we set the time threshold $TT$ to:

$$TT = \frac{1}{\mu_r} \cdot \frac{C-g}{C} \tag{3.18}$$

where $g$ stands for the number of guard channels, then the admission control condition of the proposed MRT algorithm ($MRT < TT$) is in practice that of the guard channel scheme ($BC < C - g$). Indeed, it is straightforward that for a particular value of $MRT$ and $TT \in (x-1; x)$, where $x$ is a positive integer number, the system response (admission or rejection) to new call requests will remain the same and can only change when $TT$ undergoes integer changes. Therefore, when the channel holding time is exponentially distributed the guard channel scheme can be seen as a particular case of the MRT algorithm.

We overview below the main analytical results concerning the guard channel scheme because it was demonstrated above that for classical exponential hypotheses, the guard channel scheme is a particular case of the MRT scheme. We consider a pure performance model of a single, isolated cell as noted in Section 3.1 described by the teletraffic variables defined there as well. There are $g$ number of guard channels in the system.

Let $BC(t)$ denote the number of busy channels in the system at time $t$. Then, $\{BC(t), t \geq 0\}$ is a birth-and-death proces, which can be solved mathematically. The state-transition-rate diagram of the guard channel scheme is shown on Fig. 3.6. We adopt the hypothesis for n.e.d. channel holding time and Poisson arrivals as noted previously. Let $P_j$ denote the steady-state probability that $j$ from a total of $C$ channels are busy in

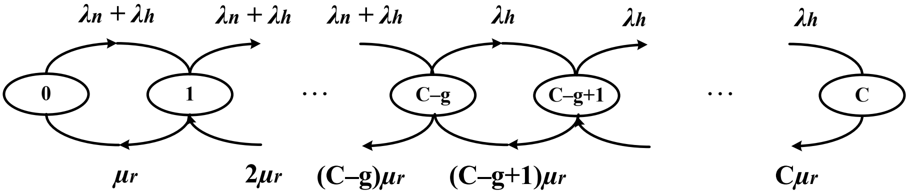


Figure 3.6: State-rate-transition diagram for the guard channel scheme, which is a particular case of the MRT algorithm when the channel holding time is exponentially distributed

the cell. This probability is easily obtained using the state-transition-rate diagram and is given by:

$$P_j = \begin{cases} \frac{(\lambda_n+\lambda_h)^j}{j!\mu_r^j} P_0 & \text{for} \quad 1 \le j \le C-g \\ \frac{(\lambda_n+\lambda_h)^{C-g}\lambda_h^{j-(C-g)}}{j!\mu_r^j} P_0 & \text{for} \quad C-g+1 \le j \le C \end{cases}$$

where

$$P_0 = \left[ 1 + \sum_{j=0}^{C-g} \frac{(\lambda_n + \lambda_h)^j}{j!\mu_r^j} + \sum_{j=C-g+1}^{C} \frac{(\lambda_n + \lambda_h)^{C-g} \lambda_h^{j-(C-g)}}{j!\mu_r^j} \right]^{-1}$$

and

$$\sum_{j=0}^{C} P_j = 1$$

The described system model has the *PASTA* property (time averages and call averages are identical) therefore, the blocking probability of new calls ($Pb$) and dropping probability of handoff calls ($Pd$) are given by:

$$Pb = \sum_{j=C-g}^{C} P_j$$

$$Pd = P_C$$

Note that computationally efficient recursive formulae for the loss probabilities of new and handoff calls for the guard channel scheme are developed by Haring *et. al* [46] and these can be applied to the MRT scheme when classical exponential assumptions hold.

## 3.6 Simulation environment

### 3.6.1 Simulator's design

We developed a pure performance simulation model for system evaluation under non-trivial conditions because of the complexity of the problem of analysing the MRT scheme under non-exponential channel holding time distributions. The simulation model incorporates the assumptions described in Section 3.1 about the cellular network. As noted there, for these common assumptions it sufficies to model and simulate the performance of one cell. We modelled the cell as a queueing system in which the radio resources assigned to the BS are the servers, while the calls (new or handoff) compose the arrival process. The channel holding time is equivalent to the service time in the queueing system. Call class differentiation (distinction between new and handoff calls) is made at the time the call request is received. A block-diagram of the developed simulation tool is included in the Appendix. We briefly comment below on its structure and functionality.

We used Omnet++ [51], which is a modular discrete event network simulator, as a simulation environment because it provides useful modules such as probability distribution functions that eased our development work. The teletraffic performance model was implemented by designing four modules. *Traffic generator* was created for generating new and handoff calls with a user-defined distribution and mean. We deveoped a *Dispatcher* module for the purposes of incorporating different admission control strategies. We implemented in the *dispatcher* the MRT algorithm but also the guard channel scheme: the time threshold $TT$ and the fixed number of channels $g$ reserved for handoff calls is a user-defined parameter[8]. We created a *Server* module as well, which modells a single resource (channel) in the system (BS). The *Statistics* block was designed to collect the output of the simulation runs[9].

We *validated* the simulator through analytical results. First, we considered a teletraffic system without a handoff prioritisation policy; that is, the non-prioritised scheme (NPS), also commonly known as complete sharing (CS) strategy. The NPS accepts all calls without differentiation given there are enough free resources in the system to serve the arriving call requests. We modelled the complete sharing policy as a special case of the guard channel scheme by setting the number of guard channels equal to zero. The NPS can be considered as well a special case of the MRT scheme when $TT \rightarrow \infty$ (in practice when $TT$ is long enough; its particular value depends on the channel holding distribution among others as

[8] The guard channels are accessed after the common pool of channels is exhausted. By common and guard pools of channels we refer to amount of resources rather than to particular channels.

[9] See next section for metrics stored in this module.

explained later in the thesis). New calls as well as handoff calls were originated according to a Poisson process. Each accepted call was assigned an exponentially distributed channel holding time with (the same) user-defined mean. The described system can be modelled through a continuous time Markov chain, for which product-form solutions are readily obtained (see [54, 61]). In particular, the system can be modelled as a $M/M/m/m$ queueing system. An excellent agreement between the experimental and mathematical results was obtained. As another means for validation, we simulated the guard channel scheme for exponentially distributed service time, Poisson arrival traffic, and different traffic loads. As noted in Section 3.5, analytical results are readily obtained for the guard channel scheme for such trivial conditions. The comparsion of the simulation and analytical results showed excellent match as well.

### 3.6.2 Simulator's parameters

#### Input parameters

The input, to the developed simulation tool, parameters are those described in Section 3.1 namely, the total number of channels, the distribution and mean of the new and handoff arrival processes, channel holding time distribution and mean.

#### Output parameters

The metrics that the simulation program gives as output are:

- $N_s^n$ – number of served new calls
- $N_b^n$ – number of blocked new calls
- $N_s^h$ – number of served handoff calls
- $N_d^h$ – number of dropped handoff calls

### 3.6.3 Performance metrics

The performance metrics of interest—blocking and dropping probabilities as well as carried traffic—are calculated from the output as follows:

- $Pb$ – probability of blocking a new call request

  It is given by:

  $$Pb = \frac{N_b^n}{N_s^n + N_b^n} \tag{3.19}$$

- $Pd$ – probability of dropping a handoff call request

  The dropping probability is a measure of the frequency with which handoff calls are dropped (not served) at a cell and is given by:

  $$Pd = \frac{N_d^h}{N_s^h + N_d^h} \tag{3.20}$$

- $Pft$ – probability of forced termination of a handoff call

  The forced call termination probability $Pft$ is the probability that the call is interrupted because one of the handoffs that the ongoing call requieres is dropped due to insufficient resources. Note that the dropping probability $Pd$ is the probability of rejecting a single handoff request. That is, $Pft$ is a measure that can be perceived in the user plane and therefore, more meaningful than $Pd$ for evaluating system-level QoS.

  The forced call termination probability can be readily found by noting that the forced call termination is the complementary event of the voluntary call finalisation. We denote with $Pc$ the probability that the handoff call will be accepted (that is, will not be dropped) in each of the visited cells, i.e., $Pc = (1 - Pd_1)(1 - Pd_2)\cdots(1 - Pd_\alpha)$, where $Pd_i$ is the dropping probability experienced in the $i$-th cell and $\alpha$ is the total number of visited cells. Because it was assumed that the system is homogeneous, that is, all cells experience the same traffic load, have the same capacity, etc., $Pd_i = Pd$ and the resultant $Pc = (1 - Pd)^\alpha$. Hence, applying elementary probability theory, the forced call termination probability $Pft$ can be determined through $Pd$ and the average number of handoff requests (average number of visited cells $\alpha$). In particular, it is given by:

$$Pft = 1 - (1 - Pd)^\alpha \tag{3.21}$$

Note that according to the ITU-T Recommendation E.771 [1], the $Pd$ should take small values, particularly, less than $5 \cdot 10^{-3}$. For such small values the forced call termination probability is proportional to the dropping probability, which is in agreement with [66, 87].

- $A_c$ – carried traffic

  The carried traffic can be computed through the input and output parameters of the simulator and is given by [11]:

$$A_c = A(1 - \frac{Pb + Pft}{\alpha + 1}) \tag{3.22}$$

The carried traffic $A_c$ is a measure of the efficiency of the system utilisation and therefore is of much interest to network operators, whereas the $Pft$ is a QoS measure associated with the user satisfaction from the service offered by the network (i.e., capability of the network to provide uninterrupted quality service to the subscribers on the move).

## 3.7 Performance evaluation of the MRT scheme for general CHT distribution

### 3.7.1 Scenarios of the experiments

The scenarios were intended to be realistic therefore we used reported empirical data for the average number of handoffs per call ($\alpha$) and call duration

Table 3.1: Simulated scenarios

| $A$<br>$[Erl]$ | $C$ | $\alpha$ | $1/\mu$<br>$[s.]$ | $SCV$ |
|---|---|---|---|---|
| 4.5; 6.5 | 10 | 2 | 40 | 1/3; 10 |

$(1/\mu)$. The mean channel holding time $(1/\mu_r)$ was determined according to (3.3). The input parameters and their feeding values used in the evaluation experiments are summarized in Table 3.1.

Even though confidence intervals are not presented, the simulation time was set long enough to ensure stationary conditions and therefore, reliable statistical data. Specifically, for all simulation runs the simulated time was $10,000$ hours, whereas the transient time interval at the beginning of each simulation run was excluded from the collected statistics.

### 3.7.2 Performance results

#### System-level QoS measures

We evaluated MRT performance for offered load of (1) 45 % (4.5 Erl) and (2) 65 % (4.5 Erl), which can be considered light and moderate load, and for Erlnag-3 and HE2b channel holding time distributions. Furthermore, although we examined a range of possible values for the time threshold (recall that $TT \in [0, \infty)$), we plot only the intervals of practical interest. The latter are limited by the blocking probability ($Pb$), for which we set an upper bound equal to 0.2.

The performance results show that the MRT policy smoothly controls the probabilities of new call blocking $Pb_i$ and forced call termination $Pft_i$ ($i = 1, 2$, where 1 indicates the case of light, and 2 – the case of moderate offered load) as a function of the time threshold $TT$ (see Fig. 3.7 and

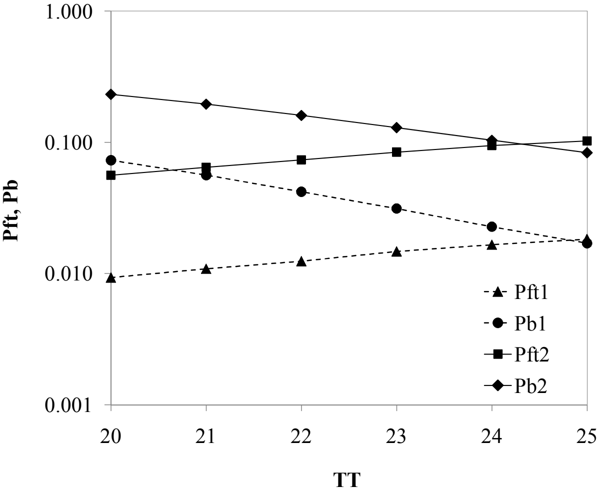


Figure 3.7: Performance of the proposed teletraffic-based scheme for (1) light (45%) and (2) moderate (65%) traffic load when the CHT is *Erlang-3* distributed

Fig. 3.8). The forced call termination $Pft_i$ is decreased at the cost of increased $Pb_i$ and vice versa. That is, the blocking and forced termination probabilities undergo reciprocal changes: when the time threshold ($TT$) is decreased, it restricts to a greater extent the flow of admitted new calls into the system and as a result the $Pb$ is increased. Simultaneously, as the time threshold is decreased, the blocking of new calls leaves on average more free channels into the system, and as a result the $Pft$ is improved.

Furthermore, we compared the MRT admission strategy to the guard channel scheme, which as noted previously has become a universal approach in evaluating admission control algorithms. It is worth mentioning here that pure loss (i.e., queueless) teletraffic systems are in general insensitive to the distribution of the holding time [54], and the guard channel scheme,

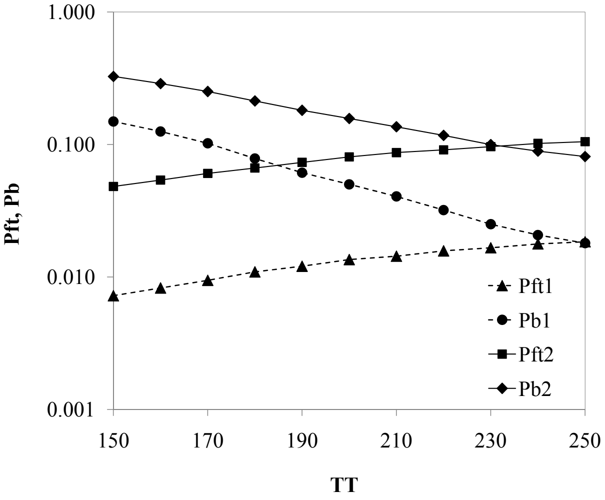


Figure 3.8: Performance of the proposed teletraffic-based scheme for (1) light (45%) and (2) moderate (65%) traffic load when the CHT is *balanced hyper-exponentially* distributed

which we denote GCS, in particular, is insensitive to the CHT distribution [110]. We validated the latter result by simulating the guard channel scheme under exponential, Erlang-3, and HE2b channel holding distributions, for which the differences in the blocking (dropping) probabilities were negligable. When the two schemes were compared, we observed that the QoS points ($Pb$-$Pft$ combinations) of the GCS can be accomplished by the MRT scheme as well (see Fig. 3.9). However, in contrast to the GCS, the MRT admission method is characterised with a continuous working interval. Consider for example, the light offered load scenario: the first three cases (number of guard channels $g$ from 1 to 3) are of practical significance, whereas for medium traffic load only the first one ($g = 1$) meets the requirement for $Pb \leq 20\,\%$ (for $g = 2$, $Pb = 0.23$). Therefore, for the GCS,

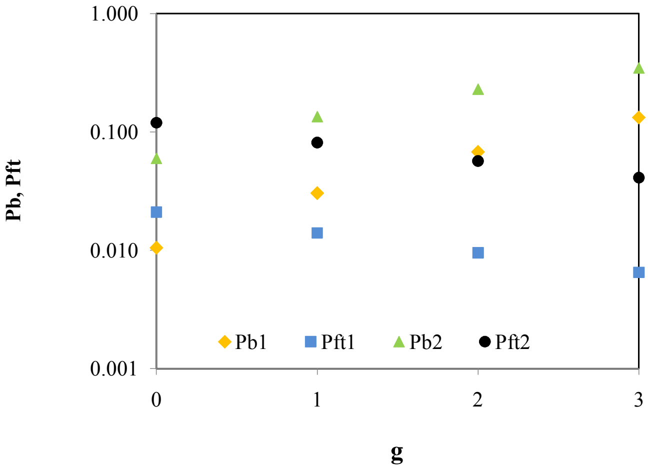


Figure 3.9: Guard channel scheme performance for (1) light (45 %) and (2) medium (65 %) offered load

the choice of the operator is very limited and restricted to a few possible working points (depending on the offered load; three or one in the cases studied). When the load offered to the system is heavier, the working interval (discrete for the GCS and continuous for the MRT scheme) becomes tighter for both admission algorithms. With the MRT scheme, however, there is still *a wider working window* compared to the traditional one, which is the major advantage of the proposed scheme in front of the GCS. The gradual transition of the blocking and forced call termination probabilities gives the operator the freedom to finely adjust the $Pb$ and $Pft$ and thus is of much value.

Various performance goals can be set and attained through the teletraffic based admission control scheme. One possible target is outlined in Table 3.2 along with the corresponding $TT$ value. It shows the case of forced call termination probability that is targeted to be five times lower than the new call blocking probability. The latter probability (i.e., $Pb$) is maintained within reasonable limits (e.g., around 5.6 % for Erlang-3 and less then 6 %

for HE2b distributed channel holding time) and the carried traffic $A_c$ is 4.4 Erl of the attainable 4.45 Erl when no handoff prioritisation scheme is implemented (i.e., for complete sharing).

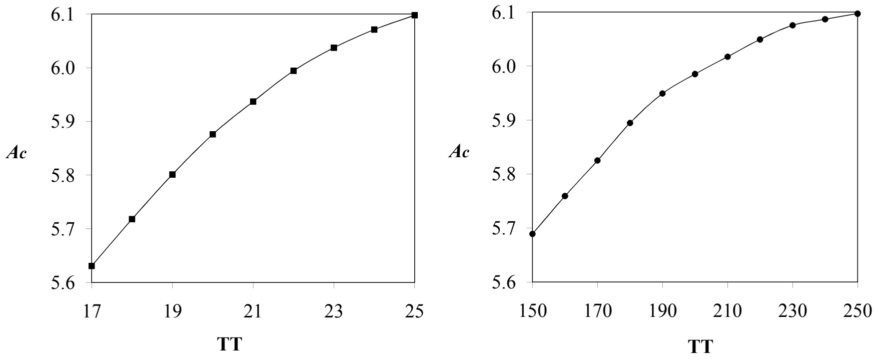
### Carried traffic

Channel efficiency utilisation is increased when the new call blocking probability, $Pb$, is decreased ($Pft$ is increased). The latter is explained by the admission into the system of a larger fraction of new call arrivals; that is, more channels are immediately allocated upon new call arrival instead of being reserved for handoff request arrivals. Moreover, the higer the carried traffic, the higher the operator's revenue. Therefore, the working point is sought according to both the target QoS and carried traffic.

The carried traffic, $A_c$, for the case of MRT implementation and moderate offered load is plotted on Fig. 3.10, whereas Table 3.3 copmrises the GCS case. For a given offered load the guard channel scheme and the MRT scheme carry the same traffic load. Importantly, however, for a given $Pb$, and (or) $Pft$ upper levels, the MRT scheme provides the operator with a wider working interval as noted earlier. In some cases, MRT scheme has higher carried traffic, which means *higher resource use efficiency* compared to the traditional one. It must be noted that the latter depends on the

Table 3.2: Performance objectives ($Pft$ five times smaller than $Pb$; $Pb$ below 6.5%) for light load $A = 4.5Erl$ and corresponding $TT$.

| CHT | $TT(5Pft = Pb)$ | $Pb(\%)$ | $Pft(\%)$ |
|---|---|---|---|
| HE-2 | 190 | 6.12 | 1.21 |
| Erlang-3 | 21 | 5.62 | 1.08 |



**(a)** Erlang-3 distributed CHT

**(b)** HE2b distributed CHT

Figure 3.10: MRT scheme: carried traffic for moderate (65%) offred traffic load and different CHT distributions

perofrmance objective. To illustrate this, we discuss below two possible performance scenarios.

Consider the case of moderate traffic load and let the operator set an upper limit on the new call blocking probability equal to 20 % (that is, $Pb \leq 0.2$) for instance. The guard channel scheme can meet this requirement only for $g = 1$ ($Pb = 0.13$, $Pft = 0.08$) because for $g = 2$ the new call blocking probability is 0.23. The carried traffic for the case of $g = 1$ is $A_c = 6.04$ Erl). For $TT = 21$ (Erlang-3), and $TT \approx 185$ (HE2b), the MRT scheme attains the following performance metrics: $Pb = 0.1954$, $Pft = 0.06445$ (Erlang-3), and $Pb = 0.19$, $Pft = 0.069$ (HE2b). The carried traffic, determined according to (3.22), is $A_c \approx 5.939$ Erl for both schemes. In other words, the operator can choose one of the two working pairs or others that lie between them when MRT scheme is implemented. Working at $Pb = 0.13$, $Pft = 0.08$ will provide the operator with higher carried traffic but higher probability of interrupting a call in progress as well, whereas working at $Pb = 0.1954$, $Pft = 0.06445$ for instance, will lead to a slightly lower carried traffic but better forced call termination probability. As explained

Table 3.3: Guard channel scheme: carried traffic under moderate offered load

| $g$ | 1 | 2 | 3 |
|---|---|---|---|
| $Ac$ | 6.03 | 5.88 | 5.58 |

earlier, the operator can seek for an acceptable tradeoff between quality of service and revenue in the case of MRT scheme.

The more natural choice for the operator though, is to set an upper limit on the forced call termination probability (or both $Pft$ and $Pb$) because operators are mainly concerned with supporting handoff (i.e., ongoing) calls continuity. Therefore, if we assume that the operator sets an upper limit on the forced call termination probabilty such that $Pft \leq 0.048$ for instance, then the number of guard channels for the guard channel scheme is set to three (when $g = 3$: $Pb = 0.346$, $Pft = 0.0405$, and $A_c = 5.663$)[10]. However, the MRT scheme attains the following working point: $Pb = 0.326$, $Pft = 0.048$, and $A_c = 5.689$ (HE2b, $TT = 150$). In other words, the MRT scheme not only meets the $Pft$ requirement as the guard channel scheme does but achieves higher carried traffic.

## 3.8 Conclusion

In this chapter we studied the applicability of the statistical profile of teletraffic variables to admission control design in mobile cellular systems. This research was motivated by (i) the scientific evidence that in general the teletraffic variables that characterise mobile cellular systems are not

[10] For $g = 2$ the forced call termination probability is $Pft = 0.0555$, i.e., above the set limit.

exponentially distributed; (ii) despite the aforesaid and to the best of the knowledge there were not published proposals that incorporate this knowledge into admission control; (iii) the scientific hypothesis that admission control schemes devised to use available statistical parameters of the main teletraffic variables can enhance system-level performance.

We carried out the work by devising admission control scheme that relies on the estimated remaining channel holding time parameter. We studied the algorithm analytically for common working hypotheses. We found that for such exponential assumptions the algorithm is identical to the classical cut-off scheme for which product-form solutions are readily obtained through continuous time Markov chain modelling. Furthermore, we studied system performance for non-trivial conditions. Due to the complexity of the problem when the channel holding time is non-exponential, we developed a simulation tool. The simulator was used for evaluation of system-level performance metrics under non-trivial conditions. Specifically, we focused on simulation experiments that relax the conventional hypothesis for exponential channel holding time but retain the assumption for Poisson arrival traffic. We considered the main system-level performance metrics that are of interest to mobile network operators, namely the probability of blocking a new call, probability of forced call termination, and carried traffic. In the following Chapter 4 we complement the performance evaluation discussion by examining the scheme for more realistic scenarios, namely high mobility, overlapping cell areas, and non-Poisson arrival process, which conditions are typical for the real mobile cellular networks.

In conclusion, through this research we demonstrated that investigations into the proposed research area can be fruitful and can lead to the design of admission control schemes that enhance system performance and user satisfaction. In particular, we showed that it can be advantageous to implement statistical estimates of system parameters, such as the channel holding time, in admission control. Through a simple decision making algorithm we expanded the discrete working interval of the classical scheme into continuous one, while we met the requirments for simplicity and fast execution of the algorithm. It should be noted that the scheme we proposed is by no means the only possible or the most efficient implementation but it achieves good overall performance and, importantly, demonstrates that statistical metrics intrinsic to pertinent teletraffic variables can be used for the development of *efficient* yet *simple* admission control policies.

# 4 System-level performance evaluation of the MRT scheme

## 4.1 Introduction

In Chapter 3 we demonstrated that the statistical nature of the teletraffic variables relevant to mobile cellular networks can be exploited in the design of admission control schemes that bring improvements into system performance. We proposed a handoff prioritisation algorithm based on statistical estimates of the (remaining) channel holding time parameter. The scheme does not reserve resources for the higher priority handoff calls in a deterministic way but in a dynamic, state-dependent fashion. The scheme admits new calls if two conditions are concurrently satisfied: (1) there are enough resources to serve the call *and* (2) a busy channel is expected to be freed within a specified time interval. Ongoing calls arriving from neighbouring cells (i.e., handoff calls) are admitted provided there are free resources in the system.

The scheme, which we named MRT after the admission control condition, was studied in Chapter 3 for queueless conditions and Poisson arrival traffic. That performance evaluation study demonstarted the advantages of the scheme over the classical one and gave us a first flavour of its behaviour. Following the methodology that we developed for studying the MRT scheme (described in Chapter 3, Section 3.4), in this chapter we focus on more realistic scenarios. The goal is to evaluate system performance for conditions typical for live mobile networks (see Section 4.1.1). We study the impact of important for the practice network, mobility, and teletraffic settings such as overlapping areas, high mobility, and non-Poisson handoff arrival process on system performance as discussed later in Section 4.3 and Section 4.4. Previously, in Section 4.2, we shortly explain the extension that we made to the simulator developed and reported on in Chapter 3. The simulator was enhanced to facilitate performance evaluation of systems with overlapping areas. Finally, in Section 4.5, we summarise our findings and conclusions.

### 4.1.1 Performance evaluation considerations

We consicely discuss below what motivated us to set each of the examined scenarios.

#### Configuration scenario I

#### Goal and motivation

This study was motivated by the mobility conditions in and layout of real and future mobile cellular networks, namely high handoff rate and (large) handoff (or overlapping) areas. The goal is to evaluate the effect that these conditions have on system performance.

*High mobility* is principally observed when (i) the perimeter of the cells is small, (ii) the velocity of users is high, or (iii) when both the aforementioned phenomena are present in the network. In urban scenarios the service coverage area of a base station tends to be smaller for higher traffic capacity [44]. Note that this is also the case of the (future) broadband mobile communication networks [90]. As a result of the small cell areas and (or) high velocity as the case might be, higher handoff rates are generated; that is, a higher mobility factor, $\alpha$, is observed. Specifically, the recent measurement data, based on extensive live traffic measurements from two commercial and mature WCDMA networks in European and Asian cities, reported by Simonsson and Lundborg [90] shows that the cell change rate can be very high. In particular, it is about 2.5 per connected minute (that is, per minute and mobile), or in other words, one cell change every 24 s. Moreover, Simonsson and Lungborg underline that as a result of the high handoff rate, the channel holding time is shorter (less than 24 s.) compared to filed studies of earlier cellular networks, according to which the occupancy time is about 40 s. (see [14]). The authors also indicate that a mean channel holding time of 20 s. is a reasonable assumption. A similar observation (handoff every 20 s.) based on field measurement data from a Finnish mobile cellular network was made by H. Holma and A. Toskala. Another relevant statistic reported by Simonsson and Lundborg [90] is that in both Eropean and Asian networks there is around 1 % of calls with more than 8 cell changes per user per minute [90]. Note, that this handoff rate might be due not only to small cell size and high user velocity but to large handoff areas as well (that is, switching to a new base station due to a better radio link). The authors, though, do not elaborate on the factors that yield such cell change rates.

In Chapter 3 we made a simplifying assumption that the handoff call is lost if all channels are busy upon arrival. However, real cellular networks feature contiguous coverage so that on one hand coverage gaps are eliminated and on the other hand continuous, uninterrupted service to the mobile users can be enabled. Often, due to the presence of overlapping areas, handoff calls are not immediately dropped but are allowed to wait for a channel assignment if all resources in the target base station are currently blocked (i.e., busy). According to Grillo *et al.* [44] another reason for the existing (large) *overlapping* (*handoff*) areas is the need for providing “adequate in-building and in-vehicle services”. The maximum time that the call can wait for a channel to be freed in the target base station is determined by the time that the call spends in the handoff area. A system with (large) overlapping areas naturally lends itself to a queueing model, which can be used for analytical and performance evaluation purposes. The maximum queueing time is determined as impatience time (i.e., the time spent in the handoff area before the mobile user loses service coverage). Note that in practice, the handoff calls are not put into a queue at the target cell but consecutive reattempts for obtaining a free channel triggered by the network are made until (i) a radio resource is assigned to the call, (ii) the mobile user moves out of the handoff area covered by the serving and target cells, or (iii) the call is completed while the mobile user is in the handoff area.

### Configuration scenario II

### Goal and motivation

The goal of the research, to be reported on shortly, is to examine the effect of the arrival process on system performance and dimensioning when admission control is incorporated into the system (i.e., base station). As in the first configuration scenario we model real conditions typical for urban scenarios: high mobility and overlapping areas. This research was motivated by analytical, simulation, and field measurements that demonstrate the non-Poisson nature of the aggregate arrival process in mobile cellular networks (see chapters 2 and 3). Recall that the arrival traffic can be approximated by a Poisson process when the population that generates the calls is large, which condition is accomplished in practice by the new traffic flow. Hence, the assumption that new call arrival traffic is Poissonian is widely accepted as realistic. However, the handoff calls arriving at a base station originate from a total number of mobile users restricted to the number of calls in progress served in neighbouring cells. In addition, the traffic profile of calls that have been served by more than one base station significantly diverges from that of the newly originated calls. It is known that smooth traffic experiences less blocking than *pure chance traffic type one* (PCT-I)[1] [54] but our goal is to obtain a quantitative measurement of handoff performance for realistic conditions when the proposed teletraffic-based scheme is implemented and, importantly, bring evidence on the significance of using accurate statistical data for designing and tuning admission control algorithms.

[1] Iversen [54] calls pure chance traffic of type I, traffic that corresponds to Poisson arrival process and exponentially distributed service times.

## Literature

Several other publications (see [40, 74] and references therein, as well as [30, 78]) have explored the effect of non-Poisson arrivals on system performance. The research conducted by Rajaratnam and Takawira [78], Dharmaraja *et. al* [30], and Feldmann [40] are the most closely related to our study. Common to Feldmann's and our work is that both investigations were motivated by empirical data that demonstrates the non-exponential nature of various traffic variables and that it is shown that the accurate modelling of teletraffic variables' profile is an important element in the design of call admission control algorithms. However, Feldmann [40] investigates *wired* networks that use a *complete sharing* scheme (i.e., all calls are admitted, provided there are available resources) and *bursty* traffic, and examines how call queuing can introduce an improvement in network performance. In our research, the impact of non-exponential inter-arrival times on *handoff performance* in a *wireless* cellular network with *admission control* and handoff queueing is investigated, with the aim of demonstrating the impact of the arrival process on system tuning. Feldmann shows that network performance does depend on teletraffic parameters, and this general conclusion is also supported by our research. However, the systems are examined for different input parameters (smooth instead of bursty traffic), which naturally leads to different output results (lower dropping probability instead of higher blocking probabilities).

Furthermore, our focus is mainly on the MRT system-level performance, although general conclusions are also drawn. Importantly, our study affirmed that system performance under non-Poisson arrival traffic conditions depends on the distribution of channel holding time, whereas for n.e.d. inter-arrival times blocking probabilities are independent of channel occupancy times. Similar result was previously obtained by Rajaratnam and Takiwira [78] for the guard channel scheme (see Chapter 3, Section 3.1.2). We examine the performance of the MRT scheme and compare it to the guard channel with queue scheme, which we denote GCQ scheme. The GCQ scheme is an extension to the guard channel scheme that allows handoff calls to wait in a queue[2]. Recall that the classical cut-off concept is

[2] As noted earlier, the queue is used in analytical and simulation studies to model the

commonly considered a reference case in performance evaluation studies. Note that the cut-off concept has been widely studied because of its high practical vaue but always under queueless conditions and (or) Poisson arrival traffic assumption[3]. To the best of the knowledge the guard channel with queue scheme is examined here for non-exponentially distributed inter-arrival times for the first time.

### Common considerations for scenarios I and II

It is worth recalling that we focus on two channel holding time cases based on the squared coefficient of variation: the Erlang-3 and balanced hyper-exponential distributions. These, as already explained in Chapter 3, exhibit contrasting behaviour, which facilitates a comprehensive study of the MRT scheme. When examined for other distributions with a different $SCV$ we expect the MRT scheme to produce the same qualitative output; that is, to perform similarly irrespective of the particular case study.

## 4.1.2 Contributions

In the previous section we set the context for the investigations that we will report on later in this chapter. In the light of the preceeding discussion our *overall contribution* is that we evaluate system performance for *conditions typical for real* and *future mobile cellular systems* and derive *useful general guidelines* for system and admission control parameterisation. In particular:

- We evaluate system performance by simulating the *high mobility* and (large) *overlapping areas* present in mobile cellular networks. Our

presence of overlapping areas in real mobile cellular networks.

[3] Xhafa and Tonguz [110, 111], for example, investigate both the guard channel and guard channel with queue schemes assuming general log-normal channel holding time distributions but Poisson arrivals. Rajaratnam and Takawira [78] and Dharmaraja *et. al* [30] study the guard channel scheme for non-trivial arrival and holding time distributions but queueless conditions.

focus is on the MRT scheme but the results also suggest some general trends that are largely independent of the implemented admission control scheme. We analyse the *mobility degradation* phenomenon for the explained in Section 4.1.1 handoff rate and cellular layout scenarios. The results establish the need for admission control implementation and show the capability of the MRT scheme to alleviate the effects of high handoff rate. Furthermore, we derive a relevant recommendation for mobile cellular network planning when high offered load and considerable mobility are evidenced. In particular, the simulation resuts show that large overlapping areas can improve handoff performance without appreciable deterioration of new call blocking probabilities, which result is important for the practice. Large overlappling areas can be found in current wireless networks due to existing radio regulations but it is important, especially for network palnning in high density scenarios (small cells, high mobility), typical for the future broadband mobile networks, that handoff areas are intentionally incorporated for system performance improvement; moreover, the perimeter of these areas should be determined based on the offered traffic and mobility conditions among other factors.

- We evaluate system performance for *non-Poisson arrival traffic* in addition to realistic load and layout conditions. That is, as in configuration scenario I system performance is assessed for high handoff rates and large handoff areas but the conventional assumtion for Poisson offered aggregate traffic is relaxed. We examine system performance for the MRT and GCQ schemes. The simulation results show that system performance is highly dependent on the channel occupancy time distribution when the arrival traffic is smooth as contrasted to the case when the aggregate arrival traffic is Poisson. Furthermore, our results affirm that the Poisson hypothesis can yield overprotection of handoff calls and thereby, underuse of system resources. The latter is translated into loss of revenue for the network operator.

## 4.2 Simulator enhancement

We enhanced the simulator described in Chapter 3 to model the existance of overlapping areas in the considered system (see Fig. 4.1). To model users' impatience time in the handoff area we incorporated a queue into the simulation model. Specifically, we modelled a queue with infinite capacity (in practice, very large number of queueing positions was used) and first-in-first-out (FIFO) queueing discipline with early abandonment. The early abondonment refers to a possible loss of service before the target base station grants a channel to the handoff call. We modelled it by introducing a maximum waiting (or impatience) time in the queue. If the queued handoff call request does not reach the first position of the queue or attains the first position but a channel is not freed in the target base station before the maximum waiting time expires, the handoff call is dropped. Note that with this extension, the schemes incorporated into the *dispatcher* module of the simulator are enhanced from queueless to queueing ones.

Similar to the validation procedure described in Chapter 3, Section 3.6.1, we considered queueing system models that were easy to solve mathematically as a means for assessing the accuracy of the enhanced simulator. In particular, we used the $M/M/m$ queueing model. The validation was carried out for different offered loads as well as different overlapping areas and mobility. Again, an excellent agreement between the theoretical and simulation results was observed. Furthermore, we modelled two simulation stopping conditions during the simulation runs: number of generated call requests and simulated time. The former condition was set to more than 5,000,000 generated calls, whereas we set the latter to 10,000 simulated hours. These stopping conditions were applied during the accuracy check and subsequently when the simulator was used for collecting performance evaluation statistics for non-memoryless conditions. The transient time interval at the beginning of each simulation run was excluded from the collected statistics. The long simulation time (or the large number of simulated calls) assured that the statistical data is reliable therefore, confidence intervals are not presented for the plotted results.

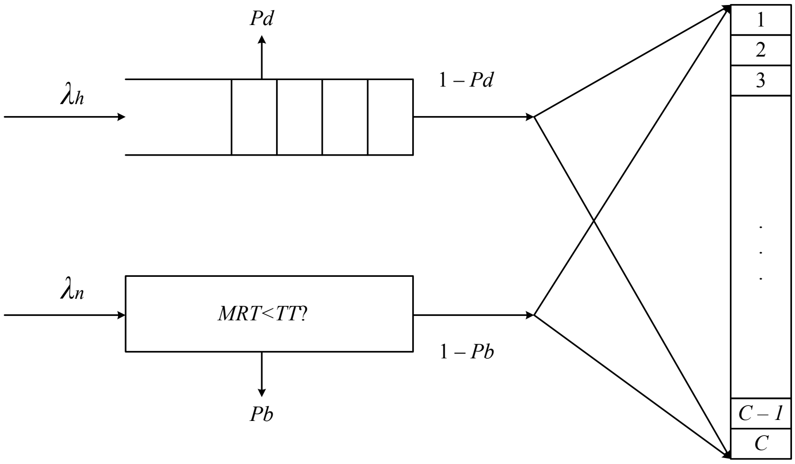


Figure 4.1: MRT scheme with queueing

## 4.3 Configuration scenario I: Poisson arrival traffic, high mobility and overlapping areas

### 4.3.1 Simulation experiments: settings

As in the previous simulation experiments (see Chapter 3), we used reported field measurement data to ensure that realistic scenarios are simulated and a posteriori analysed. In particualr, we determined the mean channel holding time considering that the normal unencumbered call duration, $1/\mu$, is approximately 120 s. [14], and that the handoff rate, $\alpha$, is high [90]. After applying (3.3) (see Chapter 3) we obtained mean channel holding time $1/\mu_r = 20$ s. and $1/\mu_r = 12$ s., which are values that agree with those suggested by the Simonsson's and Lundborg's [90] measurements. The rest of input parameters are summarised in Table 4.1. Note that we have assumed

Table 4.1: Configuration scenario I: input parameters and their values

| $A$<br>[Erl] | $C$ | $\alpha$ | $1/\mu_r$<br>[s.] | $\delta$ | *SCV* of CHT |
|---|---|---|---|---|---|
| 29; 36 | 40 | 5; 10 | $120/\alpha + 1.$ | 40; 80 | 1/3; 10 |

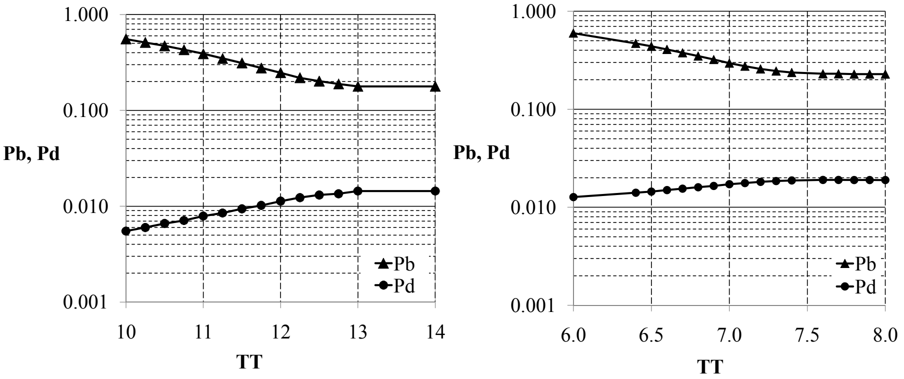
that the handoff dwell time with mean $\delta$ ($\Delta t_r$ in Fig. 2.4) is exponentially distributed. The mean of the impatience time, $\delta$, is determined as a persentage of the mean channel holding time (i.e., $\delta = x\% \, 1/\mu_r$): the higher the fraction, $x$, the larger the handoff queueing time or respectively overlapping area. The rest of the nomenclature is the same as the one specified in Section 3.1.1.

Offered traffic load of 72.5 % and 90 %, which can be considered medium and heavy load respectively, were simulated. These values were chosen taking into account that in a teletraffic loss system with full accessibility (Erlang B), system capacity of $C = 40$ total number of channels, and offered traffic $A = 29$ Erl and $A = 36$ Erl, the blocking probabilities are 1 % and 6.5 % respectively.

### 4.3.2 Performance results

Similar to the results that we obtained when the MRT scheme was studied for conventional exponential arrival assumptions and simple teletraffic conditions, the studied algorithm produced a smooth transition between $Pb$ and $Pft$ probabilities for the conditions listed in Table 4.1. As noted earlier, the new call blocking and forced call termination probabilities are mutually dependent and undergo reciprocal changes because the two types of calls (new and handoff) have a common mean channel holding time ($1/\mu_r$) and a call always occupies the same number of resources (one channel) once accepted.

Below, we examine the effect of different parameters on system performance.



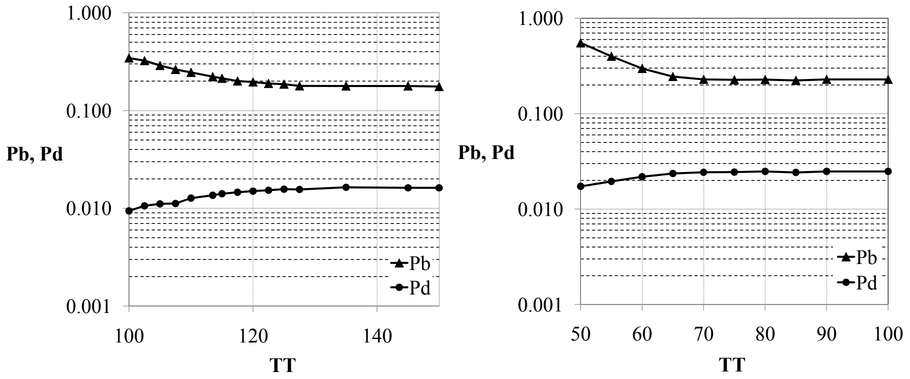
**(a)** mobility factor $\alpha = 5$ **(b)** mobility factor $\alpha = 10$

Figure 4.2: Impact of mobility: system performance with the MRT scheme for Erlang-3 channel holding time distribution, heavy load ($A$ = 36 Erl), and mean handoff dwell time $\delta = 80\%$ of the mean channel holding time

### Impact of mobility

To analyse the effect of mobility on system performance we kept the traffic load and handoff area (maximum queueing time or impatience time) constant. Recall that the arrival intensity is defined by the mobility factor $\alpha$. Therefore, for $\alpha = 10$ for example, the arrival intensity of handoff calls, $\lambda_h$, will be ten times greater than that of new calls, $\lambda_n$.

Fig. 4.2 and Fig. 4.3 show system-level performance for $A$ = 36 Erl offered load and $\delta = 80\% \, \mu_r^{-1}$. The plotted results show that high mobility (high handoff rate) has an adverse effect on both new call blocking $Pb$ and handoff call dropping $Pd$ probabilities. This performance degradation (i.e. a lower traffic capacity for the same targeted performance level) due to mobility is known as *mobility cost* [41]. It can be obserevd as well that whereas the blocking probability increases only slightly, the dropping probability increases considerably when the mobility factor $\alpha$ is rised from 5 to 10. To clarify this observation, consider first the effect of mobility on a pure



**(a)** mobility factor $\alpha = 5$

**(b)** mobility factor $\alpha = 10$

Figure 4.3: Impact of mobility: system performance with the MRT scheme for HE2b channel holding time distribution, heavy load ($A = 36$ Erl), and mean handoff dwell time $\delta = 80\%$ of the mean channel holding time

loss system. In such systems, if the offered load is constant, $Pb$ and $Pd$ will not be affected by $\alpha$, because they depend only on $A$ and not on $\lambda$ and $\mu_r$ individually [54]. In a delay system in which new calls are blocked if no channels are available and no handoff traffic is lost (i.e., the impatience time of the handoff calls is infinite; consider for instance the $M/M/m$ system), if the offered load is kept constant but traffic intensities are varied, the new call blocking probability $Pb$ will increase if the arrival intensity of handoff calls $\lambda_h$ is increased because handoff traffic that finds all channels busy is not discarded but kept until served. In addition, handoffs are treated preferentially, so their share of channels will increase at higher values of $\lambda_h$. In the simulated model, handoff calls are queued but eventually dropped if their maximum waiting time expires before a free channel is allocated to them. Therefore, the impact of handoff queueing on $Pb$ is lighter compared to the case with infinite queue and infinite impatience time. When the $\alpha = 5$ and $\alpha = 10$ scenarios are compared, it should be taken into account that both $\lambda$ and $\mu_r^{-1}$ change to maintain the offered load $A$ constant (for

Table 4.2: System-level performance comparison between the case when no admission control criterion is used and when such is implemented in the base station

| AC scheme | AC condition | $Pb$ | $Pd$ | $Pft$ |
|---|---|---|---|---|
| NPS | $TT \to \infty$, $g = 0$ | 0.1781 | 0.0144 | 0.06996 |
| MRT/GCQ | $TT \approx 12.5$, $g = 1$ | 0.2191 | 0.0125 | 0.06096 |
| MRT/GCQ | $TT \approx 11.8$, $g = 2$ | 0.2600 | 0.0108 | 0.05285 |
| MRT/GCQ | $TT \approx 11.55$, $g = 3$ | 0.3071 | 0.0093 | 0.04564 |

instance, $\mu_r^{-1}$ is 20 s. and 10.91 s. for the two examined mobility conditions respectively). In addition, as the maximum time spent in the handoff area is calculated as a percentage of the mean channel holding time, it is shorter for the latter case. Therefore, when the mobility factor is $\alpha = 10$ (i.e., high), the channel occupancy times will be shorter but the arrival intensity of handoff calls will increase and the residence time in the handoff area will fall. Consequently, the probability of dropping a handoff call grows as the mobility increases. It is noteworthy that we have plotted the handoff dropping probability $Pd$ instead of the forced call termination probability $Pft$. Note that for small values of $Pd$, $Pft$ can be approximated by applying a linear function of the dropping probability with a coefficient $\alpha$; that is, $Pft \approx \alpha Pd$ [11, 66, 87]. The probability of forced call termination grows steeply at higher mobilities because more cell changes are experienced per call on average, which increases the probability that one of the handoff requests required by the call will be dropped.

The same reasoning holds for other levels of mobility (e.g., $\alpha = 6 - 9$), but we focused mainly on moderate and worst-case scenarios (very high mobility) because, as stated above, these are typical conditions for urban mobile cellular networks. Moreover, the observed phenomena are clearly manifested in the case of high mobility.

To evaluate the improvement in handoff performance when admission control is incorporated into the system consider the results in Table 4.2, which were obtained for offered load $A = 36$ Erl, $\alpha = 5$, and $\delta = 80\,\%\,\mu_r^{-1}$. The table comprises some examplifying values for both MRT and GCQ admission control schemes and for the case when no admission criterion is used (i.e., for the non-priority scheme, which is the case when $g = 0$, equivalently, $TT \to \infty$). The results clearly show that the mobility degradation phenomenon can be combated through the incorporation of admission control strategies although at the expense of (sometimes significant) reduction of the new call arrival flow that is admitted into the system.

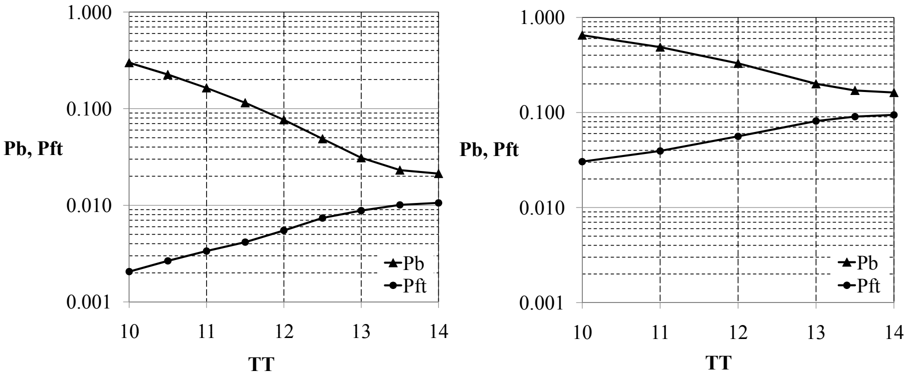
### Impact of offered traffic load

To examine the performance of the MRT scheme for different traffic conditions (medium and heavy load), the mobility and dwell time in the handoff area are kept constant at $\alpha = 5$ and $\delta = 40\%\ \mu_r^{-1}$ (which is equivalent to $\mu_r^{-1} = 120/(\alpha + 1) = 20$ s. and $\delta = 8$ s. respectively).

System perfomance is worsened for heavier offered load (see Fig. 4.4 and Fig. 4.5) because the service intensity (i.e., $1/\mu_r$) is kept the same for 29 Erl and 36 Erl offered loads, but the arrival intensities are higher in the latter case (recall that $\alpha = \lambda_h/\lambda_n$ and that for this simulation case $\alpha =$ const). Therefore, the rate of new calls blocked and calls forced to terminate are also higher; consequently, the system performance is worsened.

### Impact of handoff dwell time

Compare the results plotted in Fig. 4.2(a) and Fig. 4.3(a) with those in Fig. 4.4(b) and Fig. 4.5(b). These show the impact of the overlapping area on system performance for heavy load (i.e., $A = 36$ Erl), mobility $\alpha = 5$ but different handoff dwell time: $\delta = 80\,\%\,\mu_r^{-1}$ and $\delta = 40\,\%\,\mu_r^{-1}$ respectively. Naturally, a larger handoff area decreases the forced call termination probability, the logical consequence of which is an increase in the new call blocking probability. A longer handoff dwell time (i.e., waiting time in the queue) increases the probability that a channel will be released and a waiting handoff call request served, hence, the value of $\delta$ affects both the $Pft$

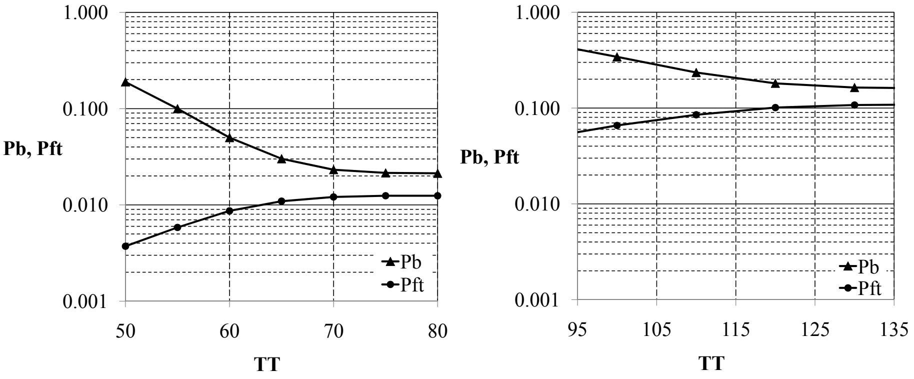


**(a)** moderate load

**(b)** heavy load

Figure 4.4: Impact of offered load: system performance with MRT scheme for Erlang-3 channel holding time distribution, mobility factor $\alpha = 5$, mean handoff dwell time $\delta = 40\,\%$ of the mean channel holding time, and different offered load

and $Pb$ probabilities. It is important to note that in all the tested scenarios, when the mean handoff dwell time $\delta$ was doubled (from $40\,\%$ to $80\,\%$ of the mean channel holding time), the resulting blocking probability was increased by a factor of about 1.05 (in practice, $Pb$ can be considered approximately constant) and the resulting forced call termination probability was decreased by a factor of about 1.7. Note that the decrease in the $Pft$ is of the same order as the decrease in the $Pd$ (recall that $Pft \approx \alpha Pd$). In other words, a larger overlapping area decreases the probability of forced termination of ongoing calls to a greater extent than it increases the probability of blocking of new calls. This result should be taken into account when designing the layout of future cellular networks because wider handoff areas improve system-level performance in presence of high mobility and (or) heavy load. Note, that the large overlapping areas in the mobile cellular networks nowadays are sometimes due to radio spectrum regulations rather than cellular planning that aims at alleviating the mobility cost.



**(a)** moderate load **(b)** heavy load

Figure 4.5: Impact of offered load: system performance with MRT scheme for HE2b channel holding time distribution, mobility factor $\alpha = 5$, mean handoff dwell time $\delta = 40\,\%$ of the mean channel holding time, and different offered load

### Impact of channel holding time distribution on queueing times

The simulation results summarised in Table 4.3 show that the mean waiting times of served and dropped handoff calls depend on the channel holding time distribution. In particular, the waiting times of calls that reach the first position in the queue and are eventually assigned a free channel and those that leave the handoff area before seizing a free channel are always shorter for Erlang-3 than for HE2b. This behaviour is due to the higher variance proper for the HE2b case. Recall that the *SCV* of the Erlang-3 is smaller than that of the HE2b distribution (1/3 and 10 respectively, see Table 4.1). Although the mean channel holding time is the same for both cases, higher dispersion around the CHT's mean produces a higher mean queueing time.

Mean queueing times of served handoff calls are longer for larger handoff areas (i.e., for $\delta = 80\,\%\,\mu_r^{-1}$ compared to $\delta = 40\,\%\,\mu_r^{-1}$), which result

Table 4.3: Mean queueing times in seconds for $A = 36$ Erl offered load and mobility $\alpha = 5$

| Type of handoff | $\delta$ | Erlang-3 | HE2b |
|---|---|---|---|
| served | 80% | 1.29 | 1.48 |
| dropped | 80% | 1.18 | 1.46 |
| served | 40% | 0.99 | 1.15 |
| dropped | 40% | 0.86 | 1.07 |

is intuitive because the maximum time that the calls are allowed to stay in the area is longer for larger handoff areas. The mean queueing time of the successfully served handoff calls in the case of $\delta = 80\,\%\,\mu_r^{-1}$ is influenced by calls that would be discarded if $\delta$ were smaller (in particular, for $\delta = 40\,\%\,\mu_r^{-1}$) because their maximum waiting time would expire before a channel is assigned. Likewise, mean queueing times of dropped handoff calls are longer for larger overlapping areas.

### Time threshold dependencies

The conclusions reported here are based on our empirical observations about the time threshold, $TT$. The variance of the CHT distribution has an impact on the $TT$. For a particular scenario and different $SCV$ of the CHT, the operational interval of $TT$ when $SCV > 1$ is usually in the order of hundreds of seconds, whereas when $SCV < 1$ it is much smaller, in the order of tens of seconds. Heavier traffic loads have a similar effect on $TT$ (higher positive values), whereas when the mean handoff dwell time is increased the operating interval of $TT$ takes on small values. A helpful clue for setting the $TT$ is the MRT performance for the exponential CHT case.

### 4.3.3 Conclusions

We studied the performance of the MRT scheme for high handoff rates and handoff areas, which conditions are typical of current and next generation wireless networks. We observed that high mobility has an adverse effect on the forced call termination probability. This result was expected because a higher handoff rate means more cell changes per active call on average and consequently increased probability of call interruption due to insufficient resources. As demonstrated, admission control can greatly alleviate the problem although at the price of reduced new call acceptance rate. Another remedy for the increased by user mobility probability of forced call interruption are the overlapping areas. Naturally, the larger the handoff area, the lower the dropping and hence, forced call termination probabilities. Importantly, we observed that large overlapping areas produce a greater relative decrease in the forced call termination probability than the relative increase in the new call blocking probability, which leads to better performance (or greater traffic capacity at a targeted performance level). These conclusions, which we derived for the MRT scheme, hold for mobile cellular systems in general. Importantly, overlapping areas should be considered by network planners, especially in metropolitan areas with heavy offered load and (or) high mobility.

## 4.4 Configuration scenario II: Non-Poisson arrival traffic, high mobility, and overlapping areas

### 4.4.1 Simulation experiments: settings

We elaborate on three scenarios. First, in scenario **A** (see Table 4.4) system performance is examined for medium and heavy load and high mobility because for such conditions admission control is relevant for providing system-level QoS (that is, $Pb$ and $Pft$ below given levels). Then, scenario **B** is configured to examine system performance sensitivity to the variations of the arrival process. Finally, in scenario **C**, we investigate the effect of the variance of the channel holding time on the performance metrics of inter-

Table 4.4: Configuration scenario II: input parameters and their values

| Scenario | $A$<br>[Erl] | $C$ | $\alpha$ | $1/\mu$<br>[s.] | $\delta$<br>$[\%\mu^{-1}]$ | $SCV$<br>of CHT | $SCV$ of<br>handoff arrivals |
|---|---|---|---|---|---|---|---|
| A | 29; 36 | 40 | 5; 10 | $120/\alpha + 1$ | 40; 80 | 1/3; 10 | 0.5 |
| B | 13 | 20 | 2.78 | 40.6 | 25 | 1/3 | 0.3; 0.7 |
| C | 40 | 50 | 2.78 | 40.6 | 25 | $1/3; 2, 3, 4$ | 0.5 |

est. In all these scenarios the handoff arrival flow is modelled as a smooth process (see Chapter 2). In particular, we model handoff inter-arrival times through Erlang-2 distribution ($SCV = 1/2$) or gamma distribution with different squared coefficient of variation but smaller than unity.

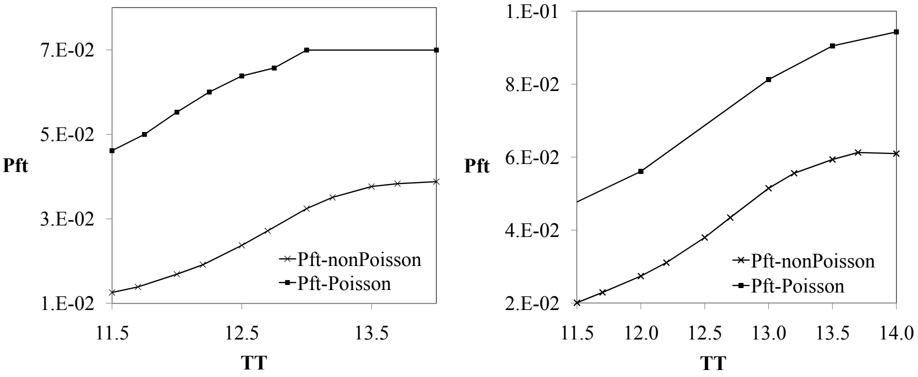
The particular choice of input values was discussed in Section 4.3.1. We complement that discussion with the motivation for the offered load generated in each of the scenarios. Scenario **A** is the same as the one studied in Section 4.3 except for the handoff arrival process, which is smooth instead of Poisson. The traffic load in scenario **B** was set according to the following consideration: a pure loss sytstem (Erlang B) with 20 channels experiences about 2 % of blocking when loaded with 13 Erl (65 % offered load). The offered load in scenario **C** is chosen to be 80 % for the same target blocking probability as in scenario **B**.

### 4.4.2 Performance results

#### Impact of arrival process distribution

The simulation results plotted in Figs. 4.6 and 4.8, and Figs. 4.7 and 4.9 show handoff performance for heavy load ($A = 90\%$) and moderate mobility rate ($\alpha = 5$) conditions.

Smooth handoff traffic leads to improved forced call termiantion probability compared to the Poisson arrival case and has a negligible effect on the



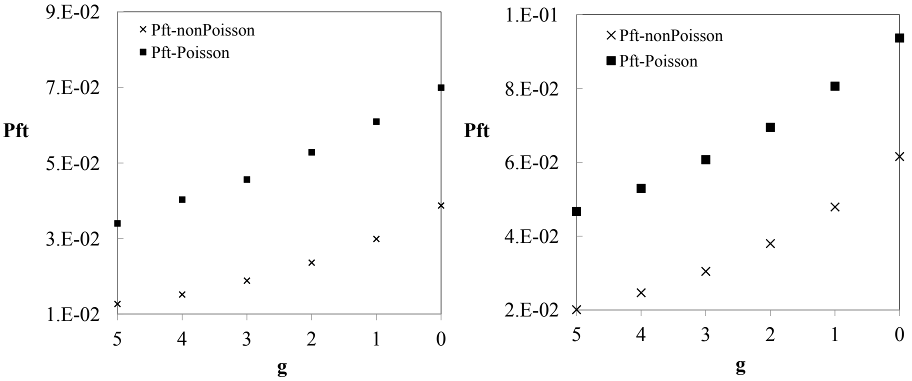
**(a)** $\delta = 80\,\%\,\mu_r^{-1}$

**(b)** $\delta = 40\,\%\,\mu_r^{-1}$

Figure 4.6: Impact of arrival process distribution: system performance with the MRT scheme for Erlang-3 channel holding time distribution, $A = 36$ Erl offered load, and mobility $\alpha = 5$ **(scenario A)**

new blocking probability. However, if the system is tuned assuming that the arrival traffic is a Poisson process whereas in practice it is smoother, the forced call termination probability will be overestimated and hence, more channels than the necessary will be reserved in the case of the GCQ scheme or more strict than the necessary time threshold will be imposed in the case of the MRT scheme. Such scheme tunning will lead to losses in both traffic and revenue. Specifically, consider Fig. 4.7 and Fig. 4.9 that show the effect of the arrival process on system tuning when the GCQ scheme is implemented in the system. Consider, for instance, a target $Pft$ of 4.5 % (i.e., $Pft^{max} \leq 0.045$), HE2b distributed channel holding time[4], and $\delta = 80\,\%\,\mu_r^{-1}$ overlapping area. Then, no more than two guard channels are needed to maintain the forced call termination probability below the upper limit (i.e., lower than 0.045). However, for the same traffic condi-

[4] System performance dependance on the channel holding time when the arrival proces is non-Poisson is discussed in a separate section following.



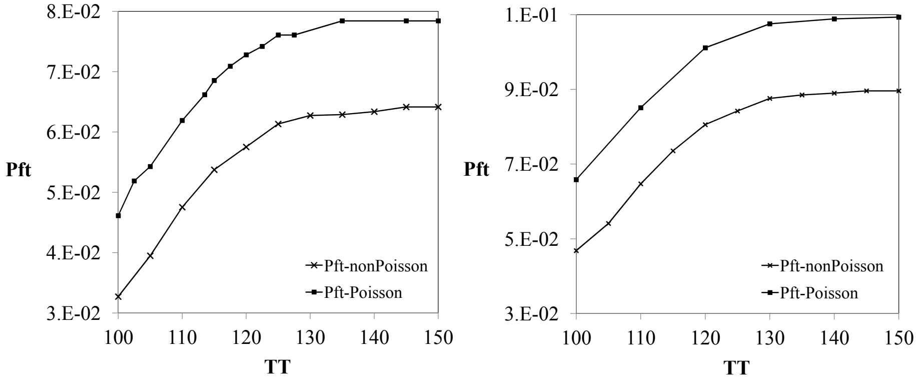
**(a)** $\delta = 80\,\%\,\mu_r^{-1}$

**(b)** $\delta = 40\,\%\,\mu_r^{-1}$

Figure 4.7: Impact of arrival process distribution: system performance with the GCQ scheme for Erlang-3 channel holding time distribution, $A = 36$ Erl offered load, and mobility $\alpha = 5$ **(scenario A)**

tions (channel holding time mean and variance, and offered traffic load) but assuming a Poisson aggregate arrival process (that is, arrival process with a squared coefficient of variation $SCV = 1$, instead of the measured $SCV < 1$), the operator will set apart no fewer than four guard channels. This would cause handoff calls to be overprotected and less new traffic to be admitted. Specifically, for the considered example, the carried traffic for $g = 2$ is $Ac = 34.21$ Erl, and for $g = 4$ is $Ac = 33.61$ Erl. As a result, the network operator will lose revenue. Similar reasoning is valid for the MRT scheme.

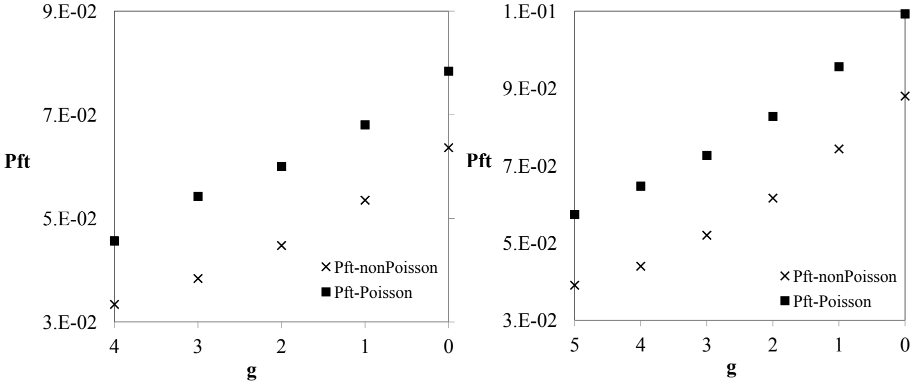
To show the advantage of the MRT scheme over the guard channel concept recall that the GCQ scheme reserves a certain number of guard channels, $g$. When the number of these channels is increased, the number of common channels to which new calls have access is decreased, so the new call blocking (forced call termination) probability increases (decreases) reciprocally. However, unlike the MRT algorithm, the GCQ scheme supports



**(a)** $\delta = 80\,\%\,\mu_r^{-1}$ **(b)** $\delta = 40\,\%\,\mu_r^{-1}$

Figure 4.8: Impact of arrival process distribution: system performance with the MRT scheme for HE2b channel holding time distribution, $A = 36$ Erl offered load, and mobility $\alpha = 5$ **(scenario A)**

a discrete working interval (i.e., $Pb$, $Pft$ pairs) because an integer number of channels is reserved. For the conditions shown in Figs. 4.8(a) and 4.9(a) if, for instance, the upper bound of new call blocking probability is 25 % (that is, $Pb^{max} \leq 0.25$) and the upper bound of forced call termination probability is 6 % (that is, $Pft^{max} \leq 0.6$), the GCQ scheme supports only one working point: $Pb = 21.2\,\%$, $Pft = 5.4\,\%$ for $g = 1$. For the same QoS requirements (i.e., $Pb^{max}$ and $Pft^{max}$), the MRT scheme supports a large number of practical $Pb$, $Pft$ combinations such as: (i) $Pb = 24.7\,\%$, $Pft = 4.8\,\%$ for $TT = 110$ s.; (ii) $Pb = 21.2\,\%$, $Pft = 5.4\,\%$ ($TT = 115$ s.); (iii) $Pb = 18.9\,\%$, $Pft = 5.8\,\%$ for $TT = 120$ s. As explained earlier, the range of the working interval is important when making a tradeoff between QoS and revenue (user requirements and operator interest) because a wider working interval provides greater tuning flexibility. Compared to the GCQ, the MRT scheme improves the restrictions imposed by the former scheme on the operator's choice.



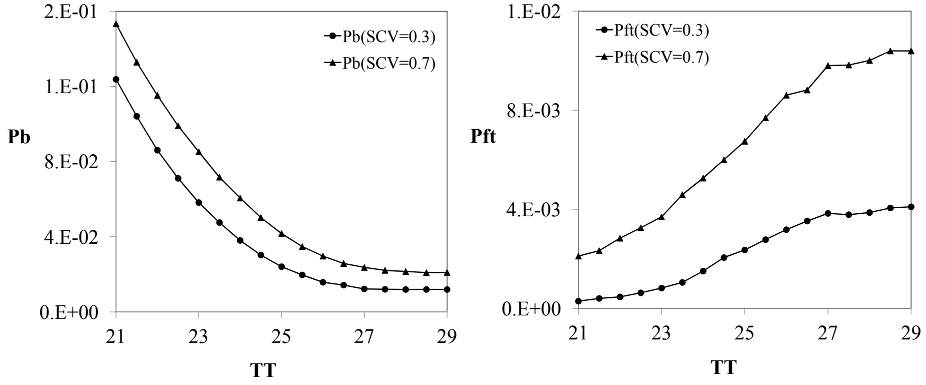
**(a)** $\delta = 80\,\%\,\mu_r^{-1}$ **(b)** $\delta = 40\,\%\,\mu_r^{-1}$

Figure 4.9: Impact of arrival process distribution: system performance with the GCQ scheme for HE2b channel holding time distribution, $A = 36$ Erl offered load, and mobility $\alpha = 5$ **(scenario A)**

### Impact of handoff arrival traffic variance

Traffic intensities as well as mean holding and mean residence times are kept constant and only the handoff arrival process is varied to determine its quantitative effect on blocking probabilities (see experiment **B**, Table 4.4). Figs. 4.10 and 4.11, show that the changes observed in the new call blocking probability and especially the forced call termination probability are consistent with the squared coefficient of variation of the arrival traffic. That is, for small values of $SCV$ fewer new calls are blocked and handoff calls dropped because the arrival process is smoother and thereby, exhibits more regular behavior compared to arrival process with higher $SCV$. In particular, when the $SCV$ is decreased from 0.7 to 0.3, the forced call termination probability is decreased by a factor of between 5 and 10. The new call blocking probability is decreased by approximately 1 %.

It is important to note that when on average more handoffs occur for a call duration, the resulting traffic is smoother. Smoother arrival traffic is



(a) $Pb$ (b) $Pft$

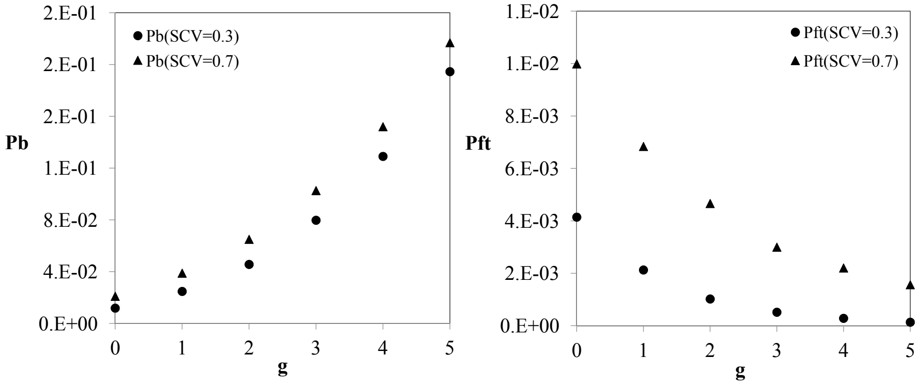
Figure 4.10: Impact of handoff traffic variance: system performance with the MRT scheme for Erlang-3 channel holding time distribution **(scenario B)**

associated with lower handoff dropping probabilities but at the same time a higher number of handoffs increases the probability of handoff forced call termination. Our results suggest the the number of handoffs prevails over the smoothness of the traffic on the resultant forced call termination probability.

### Impact of CHT's squared coefficient of variation

The results obtained from scenario **A** and plotted in Fig. 4.12 and Fig. 4.13 indicated that whereas system performance does not depend on the channel holding time distribution when the arrival process is Poisson (see Section 4.3), a dependence does exist for non-Poisson aggregate arrival traffic conditions, namely the new call blocking and forced call termination probabilities for Erlang-3 distributed channel holding time diverge significantly from the corresponding probabilities for HE2b distributed channel holding time.

Scenario **C** was set to quantify these observations, particularly, to test



(a) $\delta = 80\,\%\,\mu_r^{-1}$

(b) $\delta = 40\,\%\,\mu_r^{-1}$

Figure 4.11: Impact of handoff traffic variance: system performance with the GCQ scheme for Erlang-3 channel holding time distribution **(scenario B)**

handoff performance for smooth handoff arrival process and different squared coefficient of variations of the channel occupancy time. The differences in the $Pb$ for different $SCV$ of CHT are negligible[5], but the $Pft$ increases by a factor of up to 2 when $SCV$ is changed from 0.3 to 4 for the GCQ scheme (see Fig. 4.14). For the MRT scheme, we observed that the differences in the $Pft$ for different squared coefficients of variation of the channel holding time are of the same order of magnitude. Since for non-Poisson aggregate traffic, handoff performance depends on the channel holding time distribution as well as on the paticular arrival process, the MRT scheme as well as the GCQ scheme will require different tuning for different channel occupancy time distribution cases.

[5] Recall that we have made a common, realistic assumption that the new call offered traffic is a Poisson process. Poisson arrivals are insensitive to service time distributions beyond the mean as noted earlier.

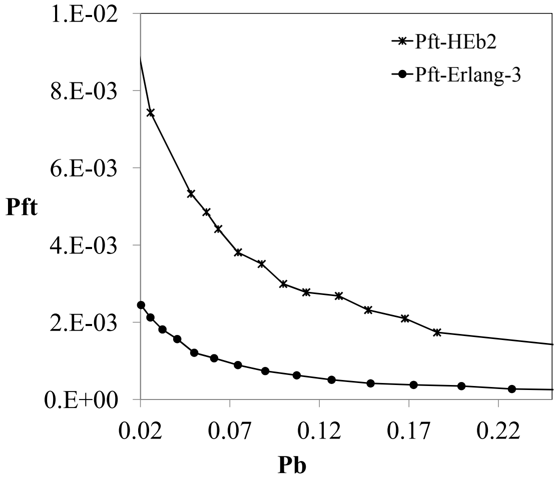
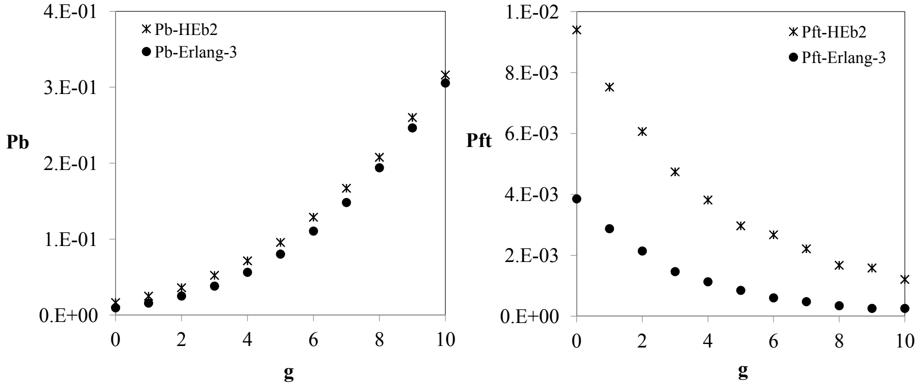


Figure 4.12: System performance dependance on the channel holding time distribution when the handoff arrival process is non-Poisson: MRT scheme case **(scenario A)**

### Impact of overlapping areas and high mobility

In Section 4.3.1 we studied the impact of real mobile cellular conditions (i.e., overlapping areas and high mobility rate) on handoff performance when Poisson arrival traffic flows are simulated. Although quantitative differences exist between that case and the case of non-Poisson handoff arrivals[6], the same qualitative conclusions can be drawn: the handoff dropping probability depends on user mobility and increases when the mobility factor $\alpha$ grows. Furthermore, the higher the offered load, the higher the new call blocking and forced call termination probabilities (see results form scenario **A**). The conclusions made about the effect of the overlapping areas on system-level QoS are also valid for non-Poisson handoff arrival traffic, namely larger handoff areas reduce the probability of forced termination of ongoing calls, $Pft$, to a greater extent than they increase the new call

[6] System performance is better for smoother handoff arrival traffic.



**(a)** $P_b$

**(b)** $P_{ft}$

Figure 4.13: System performance dependance on the channel holding time distribution when the handoff arrival process is non-Poisson: GCQ scheme case **(scenario A)**

blocking probability, $Pb$ (compare case (a) with case (b) in Figs. from 4.6 to 4.9 and Fig. 4.11).

### 4.4.3 Conclusions

We studied system performance for Poisson new and non-Poisson handoff arrival traffic flows (i.e., non-Poisson aggregate offered load) because these are realistic conditions for live mobile cellular networks. The results showed that the Poisson hypothesis for handoff offered traffic can lead to overprotection of handoff calls[7] and consequently higher penalty for the new call offered traffic. The latter has a negative effect on the efficient use of the

[7] By overprotection of handoff calls we mean that $Pft \leq Pft^{max}$ requirement can be met with lower number of guard channels in the case of the GCQ scheme or shorter time threshold in the case of the MRT with queueing scheme, when the arrival traffic is smooth. However, if the schemes are tuned for Poisson aggregate traffic, then more than the neccessary resources will be reserved.

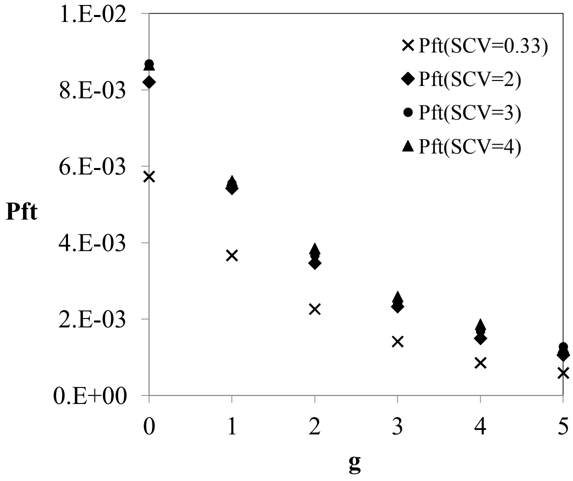


Figure 4.14: Impact of the channel holding time's squared coefficient of variation: system performance for the GCQ scheme **(scenario C)**

network resources because the actual carried traffic is less than the feasible one (that is, lower than the traffic that would be carried if handoff calls were not overprotected). In practice, more new traffic could be accepted in the system while meeting the requirements for blocking and dropping probabilities. The reduced carried traffic eventually leads to loss of revenue for the network operator. In addition, whereas the performance of the traffic-based and guard channel-based schemes does not depend on the channel holding time distribution when the arrival process is Poisson (see previous chapter as well as Section 4.4.1 for the case of MRT and [110] and [111] for GCS and GCQ respectively), a dependence does exist for non-Poisson traffic conditions. The latter result implies that scheme tuning must be done in accordance with statistical conditions.

## 4.5 Concluding remarks

In this chapter we evaluated system performance for conditions dictated by field measurements, hence, typical for live mobile cellular networks. We were interested in: (i) the effect of the overlapping areas, which are inherent in mobile systems with cellular structure; (ii) the effect of high mobility, which is typical for urban areas; and (iii) the effect of non-Poisson arrival process as is the aggregate offered traffic in practice. The system-level performance evaluation was carried out focusing on the MRT and GCQ schemes. Several important conclusions were derived.

Mobile cellular system performance is degraded for high mobility compared with lower to moderate mobility. This result is coherent with other mobility studies (see for instance [116]). Two measures can alleviate the effect of the mobility degradation phenomenon on system performance: a traditional way is the incorporation of admission control algorithms that give priority to ongoing over new call requests and therefore have the potential to greatly improve handoff performance. We demonstrated this through the traffic-based MRT scheme.

Other means for improving system performance are the overlapping areas. Large handoff areas have more noticable effect on forced call termination probability than on new call blocking probability; that is, the relative increase in $Pb$ is much lower than the relative decrease in $Pft$. More specifically, for the traffic conditions that we examined an increase of 50% of the handoff dwell time (corresponds to larger overlapping area) led to $Pb$ increased by a factor of about 1.05 and $Pft$ decreased by a factor of about 1.7. This improvement in system performance is of practical importance for network planning especially of high density areas.

The non-Poisson arrival scenario revealed some interesting results. As expected, smooth arrival traffic leads to lower blocking and dropping probabilities. Importantly, however, when the arrival traffic is not Poisson system performance depends on the first and second moments of the channel occupancy time distribution and call arrival process. This means that admission control algorithms will require different tunning conditioned on the particular teletraffic conditions.

Finally, summarising the reported results, we stress on the need for working with accurate statistics so that scheme tunning and as a result system performance, can be (close to) optimal in terms of system-level quality of service and resource use.

# 5 Validation of the MRT scheme for general channel holding time

## 5.1 Introduction

The research reported in Chapter 3 is the first to explore, from the perspective of admission control, the main statistical results derived from the vast investigations of the teletraffic variables that describe a mobile cellular network. Specifically, in Chapter 3, we proposed a new handoff priority scheme that exploits the statistical profile of the channel holding time parameter. Recall that the channel occupancy time is the time span from the instant a call (whether new or handoff) starts to use a resource that was assigned to it by its serving base station to the instant when the call releases the resource due to (whether voluntary or forced) call termination or handoff to an adjacent cell. The MRT scheme uses the distribution as well as the first and second moments of the channel holding time variable to compute the algorithm's main admission control metric, namely the remaining channel occupancy time.

In this chapter we validate the MRT scheme for more restrictive conditions than those assumed in its design. Our goal is to generalise the handoff scheme for the case when only the mean and variance of the channel holding time are known, but precise knowledge of the complete channel occupancy time distribution is not available. The problem is of practical interest because probability distribution fitting can be time- and CPU-intensive. We use an approximation of the expected remaining holding time instead of its exact value (see Section 5.2). We show that for a variety of traffic conditions (see Section 5.3) and the general gamma distribution, the approximation yields acceptable results (see Section 5.4). For a log-normal distribution, our results suggest that the algorithm can be defined based solely on the mean and variance of the channel holding time variable (see Section 5.4 and the concluding Section 5.5).

### 5.1.1 Motivation and goal

In the previous chapters 3 and 4 we analysed the proposed traffic-based handoff priority scheme under the assumption that the probability distribution of the channel holding time was known. We conjectured that the same scheme would maintain its performance characteristics when operating with only the mean and variance of this random variable. For implementation purposes, the latter scenario is of practical interest due to the following reasons. First, the numerical process of probability distribution fitting can be lengthy and heavy. Moreover, the CPU load resulting from the statistical process of finding the first two moments of the channel holding time is much lower than that of determining the whole distribution. Second, the channel holding time distribution depends on factors such as the mobile user velocity, cell size, and unencumbered call duration. As a consequence of mobility and traffic pattern variations, which can be experienced on the time scale of days or weeks, the channel holding time distribution can exhibit time and space variations. Third, the analytical expressions used in the algorithm for different probability distributions can be approximated by only one equation (see Section 5.2). Hence, if the conjecture that the algorithm keeps its functionality in such conditions, is proven to be valid, this will facilitate the practical implementation of the algorithm[1].

[1] Recall that wireless network operators choose admission control schemes based on two primary criteria: overall performance and implementation complexity of the algorithm.

### 5.1.2 Contributions

The goal of the research reported below is to test the applicability of the proposed admission control scheme when the first two moments of the channel holding time variable are known, but the remaining descriptors of its distribution are not available.

The chief contribution of our work is that we propose an approximation method for determing the main admission control metric of the MRT scheme. The proposed approximation allows probabilistic estimation of the remaining channel holding time knowing only the first and second moments but not the complete distribution of the channel holding time random variable. Naturally, the proposed approximation is not exact yet our results suggest that it can be used when only the mean and variance of the channel holding time are available. Note that we validated the approximation for log-normally and gamma distributed channel holding time.

## 5.2 Proposed approximation

### 5.2.1 Evaluation setup

We verify our conjecture about the algorithm's functionality for the following conditions:

- The aggregate traffic (new and handoff) is assumed to be a Poisson process.
- The residence time in the overlapping area follows a negative exponential distribution with mean $\delta$. Furthermore, the handoff dwell time is modelled as a queue with early departure: handoff calls that abandon the queue without having been served lead to interrupted calls.
- The hypothesis for exponentially distributed channel holding time is relaxed. The proposed approximation is tested for channel holding time distributions with behavior sparser than the exponential, namely

with a squared coefficient of variation greater than unity, which is typical of voice calls according to field trials.

### 5.2.2 Approximating the expected remaining channel holding time

The conjecture that we want to study is whether the MRT algorithm is functional when only the mean $(1/\mu_r)$ and variance (or equivalently $SCV$) of the channel holding time are available.

Recall that in Chapter 3, Section 3.2 we derived the mean remaining channel occupancy time conditioned on the elapsed time since the channel was occupied, see (3.8). As demonstrated there, the mean remaining holding time, $\overline{h}(\epsilon)$, is defined by the conditional probability density function of the residual occupancy lifetime, i.e., $\hat{f}(t|\epsilon)$; see (3.8) and (3.7) respectively. However, we want to study the case for which the probability density function that best fits the channel holding time is not available at the base station but the mean and variance of this random variable are known based on measured data. Therefore, in order to estimate the mean remaining time we have to first approximate the probability density function of the channel occupancy time. The approach we suggest is to use the balanced hyper-exponential distribution so that $\overline{h}(\epsilon)$ can be computed independently from the actual channel holding time distribution, which is assumed to be unknown. The proposed approximation of $\overline{h}(\epsilon)$ avoids the sometimes cumbersome process of distribution fitting as explained earlier. For the balanced hyper-exponential distribution, the mean remaining channel holding time is given by (3.15), which we repeat here for completeness:

$$\overline{h}(\epsilon) = \frac{1}{\mu_1}\frac{pe^{-\mu_1\epsilon}}{pe^{-\mu_1\epsilon}+(1-p)e^{-\mu_2\epsilon}} + \frac{1}{\mu_2}\frac{(1-p)e^{-\mu_2\epsilon}}{pe^{-\mu_1\epsilon}+(1-p)e^{-\mu_2\epsilon}} \tag{5.1}$$

where $\mu_1$ and $\mu_2$ are the rates of the two exponential stages of the hyper-exponential distribution, and $p$ and $(1-p)$ are the probabilities of entering the exponential stage with mean $\mu_1^{-1}$ and $\mu_2^{-1}$ respectively.

Because we use the HE2b approximation for calculating the mean remaining $\overline{h}(\epsilon)$ of a general channel holding time with first and second order metrics (i.e., $\mu_r^{-1}$ and $SCV$ are known), these moments must be fitted first to the parameters of the balanced hyper-exponential distribution. This can be done by applying the following set of equations [9]:

$$p = \frac{1}{2}\left(1 - \sqrt{\frac{SCV - 1}{SCV + 1}}\right) \tag{5.2}$$

$$\mu_1 = 2p\mu_r \tag{5.3}$$

$$\mu_2 = 2(1-p)\mu_r \tag{5.4}$$

In other words, first, the measured mean and variance of the channel holding time must be fitted to $\mu_1$, $\mu_2$, and $p$, which are the parameters of a two-stage hyper-exponential distribution. Then, the mean remaining channel holding time is estimated according to (5.1). Note that the mean remaing time $\overline{h}(\epsilon_i)$ of channel $i$ depends on three metrics that are easy to obtain as noted before. The elapsed time $\epsilon_i$ since channel $i$ has been occupied is recorded in the base station (such data is readily available because it is needed for charging purposes for example). The other two metrics, namely the mean $\mu_r^{-1}$ and variance $SCV$ of the channel occupancy parameter, are determined after a simple statistical process. Finally, the $MRT$ admission control metric is given by (3.9), which we repeat here:

$$MRT = \frac{1}{C}\sum_{i=1}^{C}\overline{h}(\epsilon_i) \tag{5.5}$$

### 5.2.3 Methodology

We use a simulation approach[2] for validating the proposed approximation. The reason for this is that we examine our hypothesis by considering a log-

[2] The system model and the simulation tool that we devised for system performance evaluation were explained in the preceding Chapters 3 and 4.

normally and gamma distributed channel holding time as is to be explained below. These distributions can not be decomposed into exponential stages and therefore lack the memoryless property of the exponential distribution. Whereas the mathematical analysis for exponentially distributed variables is simple, for log-normal and gamma it is not straightforward.

Recall that independent empirical studies in wired but also land-mobile networks use the log-normal distribution as a best fit to measured data in representing the call but also channel holding time random variables (see Chapter 2). This motivated our choice for the log-normal distribution. In addition, we wanted to study other random variables because although accurate the field results are limited to measured sites and might not reflect all the possible conditions that can be found in practice. Our concrete choice of gamma distributed random variable is backed by its versatility in representing other probability distributions.

Fig. 5.1 illustrates the remaining channel holding time versus the elapsed one for the examined log-normal and gamma, as well as for the HE2b and exponential[3] distributions, when the mean CHT is $\mu_r^{-1} = 20$ s., and the CHT squared coefficient of variation is $SCV = 5$.

To measure the accuracy of the approximation that we propose for estimating the residual channel occupancy time we proceeded as follows. First, we simulated the channel holding time according to a given distribution (either the log-normal or the gamma) but we estimated the mean remaining channel holding time $\overline{h}(\epsilon_i)$ in each busy channel $i$ according to (5.1). Second, we used the same simulation scenario (channel holding time distribution and the rest of input parameters) but we estimated the mean remaining time according to (3.8); that is, respecting the actual channel holding time distribution used for the simulations[4]. Then, the exact and approximated case were compared. Additionally, based on our previous research results, we devised another method for verification. Recall that the results reported in Chapters 3 and 4 demonstrated that the guard channel scheme is a particular case of the MRT scheme when the channel holding

[3] The mean remaining time $\overline{h}(\epsilon)$ for an exponentially distributed random variable when the elapsed time is $\epsilon$, $\epsilon \geq> 0$ equals its mean, see Chapter 3.

[4] In particular, for estimating $\overline{h}(\epsilon)$ of the log-normal and gamma distribution we numerically integrated (3.8) in Matlab.

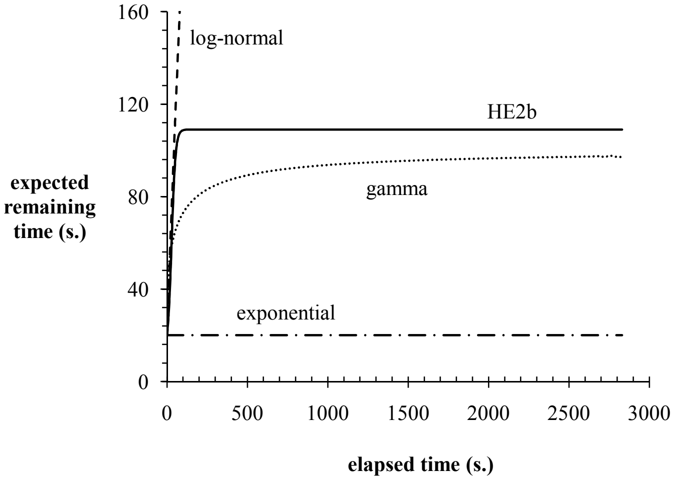


Figure 5.1: Estimated mean remaining versus elapsed channel holding time. The CHT has mean $\mu^{-1} = 20$ s. and $SCV = 5$

time is exponentially distributed. Moreover, it was shown that for other occupancy time distributions the guard channel is a discrete case of the MRT scheme. In other words, the working points of the guard channel scheme are a fraction of the working interval of the MRT scheme. Thus, we have two ways of measuring the accuracy of the proposed approximation approach. A comparison of the approximation, which we denote MRT$_{approx}$ scheme, with the actual, which we denote MRT scheme as before, or with the guard channel, shows the divergance (error) introduced by the approximation from the exact values. Further, the guard channel concept can be considered as before a reference case for system performance evaluation.

Table 5.1: Verification of the MRT scheme: experiments set-up

| $A$<br>[Erl] | $C$ | $\alpha$ | $1/\mu$<br>[s.] | $\delta$<br>[% $(\mu_r^{-1})$] | CHT's<br>SCV | CHT's<br>distribution |
|---|---|---|---|---|---|---|
| 29; 36 | 40 | 5; 10 | $120/\alpha + 1$ | 0; 40; 80 | 5; 10 | log-normal; gamma |

## 5.3 Simultaion scenarios

The input teletraffic parameters and their values, which were intended to fit realistic cases, are listed in Table 5.1. The offered traffic load $A$ was set to 29 Erl and 36 Erl taking into account that a teletraffic loss system with 40-channel-capacity (i.e., $C = 40$) and full accessibility when loaded with such traffic will experience blocking probabilities (Erlang B) equal to 1 % and 6.5 % respectively. These traffic conditions can be considered medium and heavy offered loads. Scenarios with high mobility were preferred over scenarios with low mobility in order to evaluate system performance for urban conditions (high density, consequently small cell sizes and high mean number of handoffs per call, $\alpha$). That is, we simulated conditions for which admission control implementation is essential for meeting QoS requirements. Moreover, for such conditions, performance trends are more easily observed than for other low mobility and offered load scenarios. The mean call duration was set to 120 s., which is typical for voice service in mobile telephony. Recall that the mean channel holding time, $\mu_r^{-1}$, is given by the mean call duration, $\mu^{-1}$, over the mean number of handoffs per call plus one (the cell where the call is initiated plus the number of handoffs per call), $\alpha + 1$ (see (3.3)). Handoff areas (modelled by the maximum time $\delta$ that the call can remain in the overlapping area) are also considered.

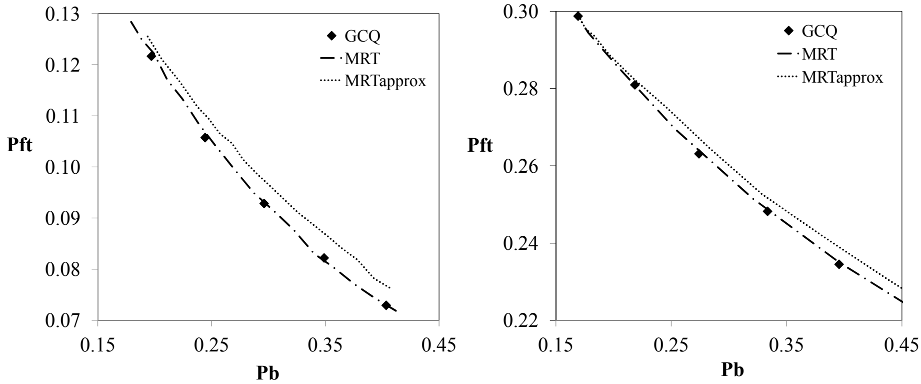
## 5.4 Performance results

### 5.4.1 Gamma-distributed CHT

The results plotted in Fig. 5.2 and Fig. 5.3 are for medium to high mobility, and offered traffic $A = 36$ Erl, which can be considered heavy load.

The results for gamma-distributed channel holding time affirm our earlier conclusion that the guard channel (with queue) scheme obtains a fraction of the working interval of the MRT (with queue) scheme. Note that network operators can set different upper system performance limits in terms of blocking (dropping) probabilities. If the maximum acceptable level of the new call blocking probability is set to less than 25 %, yet a very high value, the guard channel with queue scheme supports a very limited choice: only two possible $Pb$, $Pft$ working pairs for the examined teletraffic conditions. In comparison with the GCQ scheme, the MRT scheme, due to its continuous working interval, provides more freedom to select a working point that meets required QoS and carried traffic criteria. Furthermore, and as noted previously, the observed interrelated behavior between $Pb$ and $Pft$ agrees with intuition: restricting the admission of new traffic into the system provides for more free resources for handoff call requests, which leads to lower dropping respectively forced call termination probabilities but results in higher blocking probability.

Importantly, we measured the approximation approach (MRT$_{approx}$) divergence from the exact one (MRT): the difference for the working range of practical interest is less than 0.5 %; that is, for the same new call blocking probability $Pb$, MRT$_{approx}$ obtains forced call termination probability $Pft$, which is at most 0.5 % higher than that of the MRT scheme. Consider Fig. 5.1, the estimated mean remaining time of the HE2b distribution is longer than the estimated mean remaining time of the gamma distribution for the same $\mu^{-1}$ and $SCV$, which explains the observed differences. Moreover, when the new call acceptance ratio is further limited (i.e., for shorter time treshold $TT$), the accuracy of the approximation decreases. The weight of the error introduced by the approximation is higher for lower values of $TT$. Conversely, a larger $TT$ introduces larger tolerance to errors in the estimated remaining time.



**(a)** mobility factor $\alpha = 5$, $\mu^{-1} = 20$ s. **(b)** mobility factor $\alpha = 10$, $\mu^{-1} \approx 10.91$ s.

Figure 5.2: System performance for gamma-distributed channel holding time with mean $\mu^{-1}$ and $SCV = 5$ and for mean handoff dwell time $\delta = 40\%$ of the CHT

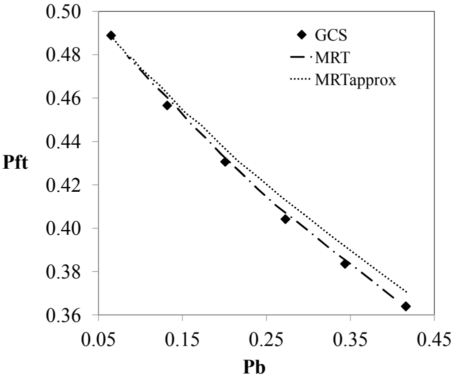


Figure 5.3: System performance for gamma-distributed channel holding time with $SCV = 5$, mobility $\alpha = 10$, and no queuing

Mobility does not have effect on the accuracy of the approximation. Note, however, that a higher handoff rate—higher number of visited cells during a call—in a homogeneous mobile cellular network yields higher forced call termination probability as shown before.

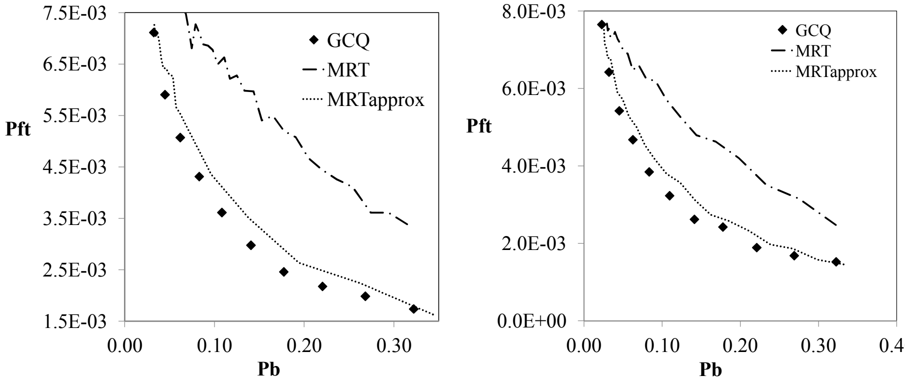
Fig. 5.3 illustrates MRT$_{approx}$ system performance when queueing is not allowed (corresponds to systems with very narrow handoff areas or to systems where repeated handoff requests are not allowed: if a handoff request is rejected the call is lost). The same qualitative conclusions can be made as for the case with queueing and only quantitative differences exist: the blocking (dropping) probabilities are prohibitively high when overlapping is insignificant and heavy traffic and high mobility conditions are present in the system.

Finally, we examined the sensitivity of the MRT$_{approx}$ scheme to the CHT variance when handoff calls are (not) queued. The MRT$_{approx}$ accuracy does not depend on the overlapping area. Moreover, our results affirmed that MRT$_{approx}$ like MRT queueless system performance is not sensitive to the channel holding time distribution: the differences in $Pb$ and $Pft$ are negligible for different $SCV$ of the channel holding time[5].

### 5.4.2 Log-normally-distributed CHT

The results illustrated in Figs. 5.4 and 5.5 are for offered traffic $A = 29$ Erl, which can be considered moderate load. The MRT scheme maintains its principal characteristic of providing a continuous working interval when a log-normally-distributed channel holding time is simulated. However, there are important quantitative differences between $Pb$, $Pft$ working points of the MRT and GCQ schemes for log-normally-distributed channel holding time. Specifically, in contrast to the gamma case, the performance of the MRT$_{approx}$ for log-normally-distributed CHT is closer to that of the GCQ scheme than is the MRT scheme: that is, MRT$_{approx}$ achieves lower blocking (dropping) probabilities of new (handoff) traffic than the exact MRT, which result was not expected. Recall that for all the experiments reported

[5] This result agrees with the performance of a pure loss system when the arrival process is Poisson, namely the system is not sensitive to moments higher than the first [54].



**(a)** CHT $SCV = 10$ **(b)** CHT $SCV = 5$

Figure 5.4: System performance for log-normally-distributed channel holding time with mean $\mu^{-1} = 20$ s., for $\delta = 80\%$ of the CHT and $\alpha = 5$

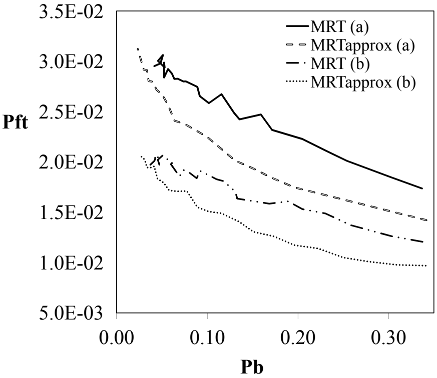


Figure 5.5: System performance for log-normally-distributed CHT with $SCV = 10$, $\alpha = 10$ and queueing time **(a)** $\delta = 40\%\,\mu^{-1}$ and **(b)** $\delta = 80\%\,\mu^{-1}$

previously the guard channel and MRT schemes attained the same working pairs. Moreover, the MRT scheme expanded the discrete working interval of the guard channel scheme to a continuous one. The results for the log-normal channel holding time, however, differ from this general trend as can be seen from the figures.

Furthermore, the differences between the exact and approximation approach (i.e., between MRT and MRT$_{approx}$) are two to three times higher than the corresponding values for a gamma-distributed channel occupancy time. This can be explained by considering Fig. 5.1, which shows the mean remaining versus the residual lifetime for HE2b in comparison with log-normally- and gamma-distributed random variables. Note that HE2b is a better approximation of the gamma than of the log-normal remaining time, $\overline{h}(\epsilon)$, therefore, the differences are larger for the log-normal case.

We tested system performance sensitivity to the variance of the channel occupancy time when queuing is allowed (see Figure 5.4). As for the gamma-distributed CHT case, MRT$_{approx}$ handoff performance for log-normally-distributed channel holding time depends on the CHT's *SCV* when handoff calls are queued. A higher coefficient of variation increases the dropping rate and consequently increased the forced call termination probability. Furthermore, the higher the variance of the channel holding time, the higher the differences between the performance of the MRT and the GCQ scheme. Queueing times (overlapping areas) have effect on the quantitative measures[6], but not on the general performance of the MRT scheme and the approximation approach (see Fig. 5.5).

## 5.5 Conclusion

In this chapter we studied the performance of the MRT scheme for the assumption that the distribution of the channel holding time is unknown but the first two moments are available. The problem is of practical interest because statistical probability distribution fitting can be time- and CPU-intensive.

[6] The longer the dwell time $\delta$ in the handoff area, the higher the probability that an ongoing call will be successfully handed off to a target cell.

We approximated the main metric—expected remaining time—by implementing balanced hyper-exponential distribution, which turned out to be a feasible approach when only the mean and variance of the channel holding time are available. In particular, the approximation yielded acceptable results for the computation of the remaining channel holding time when the channel occupancy time was gamma-distributed. The results for log-normally-distributed channel holding time showed better system performance when the approximation was used. The latter result suggests that the mean and variance might be sufficent to define the main parameter of the MRT scheme, namely the mean remaining channel holding time, for the log-normal case. Future work could investigate further on other feasible approximations of the mean residual lifetime for the conditions described above. In particular, other probability distributions such as phase-type ones can be explored. The latter work is of importance especially for the log-normal channel holdinh time case, which showed behaviour different from the expected one.

# Part II

# Mobile cellular networks with broadband access

# *6* Mobile WiMAX

## 6.1 Introduction

The mobile traffic has been exhibiting an unceasing growth since the deployment of the very first commercial mobile cellular networks. The ever increasing number of mobile subscribers[1] has experienced exponential growth even in conditions of economic downturn, as those observed during the last three-year-period, exceeding the forecasted one [3]. Moreover, the reported untethered access to the Internet in 2010 was three times the entire global Internet in 2000 [3]. At the beginning of the mobile wireless era, the evidenced teletraffic rate was primarily a result of the increasing number of mobile subscribers. Today[2] and especially in the future, the traffic growth is expected to be sustained by fundamental trends such as: (i) emergence of new services and applications, (ii) evolution of the mobile terminals and their functionality, (iii) migration from fixed to mobile communication[3],

[1] There are already 48 million people who do not have electricity at home but have mobile phones according to [3]. It is estimated that they will amount to 138 million people by 2015 [3].

[2] There are 5 billions of mobile subscribers (the world population nowadays is about 7 billions according to Wikipedia) and 2 billions of fixed Internet connections today. It is anticipated that there will be 788 million mobile-only Internet users by 2015 [3].

[3] A traffic migration from the fixed to the mobile network is already taking place in the developed countries. In Finland, for instance, this transition is at an advanced state:

(iv) increasing terminals' usage rate per person, (v) proliferation of new paradigms such as machine-to-machine (M2M) communications, (vi) rollout of mobile cellular networks[4], in addition to (vii) the increasing but finite number of subscribers.

In an effort to meet the mobile traffic growth rate and user requirements for quality of service equivalent to the supported by wire, The International Telecommunication Union – Radiocommunication Standardization Sector (ITU-R) started more than ten years ago a process towards the International Mobile Telecommunications systems (IMT) – Advanced development. At the end of 2010, Long Term Evolution-Advanced (LTE-Advanced) and IEEE 802.16m were qualified by ITU-R as IMT-Advanced systems. By incorporating several physical (PHY) layer advanced techniques these technologies are expected to provide larger capacity, higher data rates and more diverse service environment compared to the third generation (3G) cellular networks.

The main interest in Part II of the monograph is on the new context set by the recently developed mobile broadband technologies within which admission control shall be devised and executed. As the admission control investigations are inevitably subordinated to the implemented technology, we were forced to focus on a specific telecommunication standard. We considered the mobile Worldwide interoperability for Microwave Access (WiMAX) technology, which is based on the IEEE 802.16-2005 standard (also known as IEEE 802.16e) because it is representative for the mobile broadband access technologies based on the orthogonal frequency division multiple access scheme. The latter is becoming widely adopeted in telecommunication and especially wireless systems due to the improvements in capacity introduced by that access scheme. Importantly, several mobile broadband technologies

the mobile phones have substituted to a large extend the fixed terminals and younger generations have never had a land-line connection in their homes due to the mobile terminals' proliferation and the low prices supported by mobile operators.

[4] The rollout of forth generation (4G) mobile networks in the developed as well as of third generation (3G) mobile communication systems in the developing countries, which lack wired infrastructure, provides immense opportunities to mobile operators [3]. The increase in the supported speed of the network connection leads to increased usage of the network.

exploit, expand and enhance the engineering advances originally incorporated in the IEEE 802.16 family of standards, which motivated our choice.

In this Chapter 6 we provide a succinct overview of the main buliding blocks of the advanced wireless cellular systems. These are shortly explained (see Section 6.2), primarily from the angle of the research objectives set in Chapter 7 following. Then, we provide a concise summary of the WiMAX technology (see Section 6.3).

## 6.2 Building blocks

### 6.2.1 OFDMA

The high bandwidth wireless transmissions typical for the broadband communications are more prone to frequency-selective fading and as a result to signal distortion than lower bit-rate transmissions. *Orthogonal frequency division multiplexing* (OFDM) efficiently mitigates these propagation effects as explained below. Consequently, OFDM supports good resistance to multipath and operation in *non-light-of-sight* (NLOS) conditions, which are among the chief reasons for its integration in the advanced mobile broadband technologies.

OFDM is a digital multicarrier modulation. It divides the bitstream to be transmitted into several substreams. The bitrate $R_N$ of the substreams is significantly lower than the total rate $R$ of the original bitstream and so is the bandwidth $B_N$ of each individual substream compared to the total bandwidth $B$ of the channel. The substreams are sent over equally spaced frequency subchannels (often called *subcarrier channels* or *tones*), which are orthogonal to each other. The number of subchannels $N$ is chosen to make their bandwidth[5] sufficiently large or equivalently the OFDM symbol time $T_N \approx 1/B_N$ on each subchannel significantly larger than the delay spread of the channel, so that the subcarriers experience relatively flat (non-frequency-selective) fading. The *inter-symbol interference*[6] (ISI)

[5] Note that $B_N = B/N$ and $R_N \approx R/N$.

[6] Recall that inter-symbol interference occurs when successive digital symbols overlap into adjacent symbol intervals.

on each subcarrier as a result is negligable. Furthermore, the subcarriers need not be contiguous, which avoids the need for large contiguous blocks of radio spectrum for high rate communications. A serial-to-parallel converter is used and all the $N$ carrier-modulated signals are simultaneously transmitted over the OFDM symbol interval $T_N$[7].

The *orthogonal frequency division multiple access* (OFDMA), similar to OFDM, divides a fast signal into many slow signals spaced apart. The difference between the two variants of the orthogonal technology is that OFDMA can dynamically assign a subset of all the subcarriers to individual users (data regions in the WiMAX subframe, see below), which facilitates flexible allocation of resources depending on the channel state.

### 6.2.2 Link-adaptation

Wireless systems that do not adapt to the time-varying radio channel conditions require deterministic link margin to maintain acceptable performance when the channel is poor. These systems (primarily the conventional, voice-oriented mobile cellular systems discussed in Part I) are designed for the worst-case radio link scenario, which can result in (highly) inefficient use of the radio channel resources. To address this issue, the advanced mobile communication systems incorporate link-adaptation techniques. The latter allow the system to adapt to the fluctuating radio channel and hence, optimise the achieved throughput. In particular, *adaptive modulation and coding* is a spectrum-efficient technique that enables robust transmission over time-varying propagation conditions. The *hybrid automatic repeat request* when combined with AMC can further improve system performance. Another means of taking advanatage of the varying radio propagation conditions is the use of *multiple in multiple out* antenna systems. We briefly overview below the basic principles of each of these techniques.

[7] Among the several excellent sources that provide a detailed discussion on this topic as well as the techniques following are [43, 87].

### AMC

Modulation (or variation of the frequency of a carrier) is an universal means for transmitting data in communication systems at desired operating or carrier frequency. From second generation (2G) systems on, digital instead of analog modulation schemes are used, particularly, the modern mobile communication technologies incorporate *quadrature phase shift keying* (QPSK) and M-ary *quadrature amplitude modulation* (QAM). Furthermore, to increase the resilence to errors some overhead in the form of *forward error correction* (FEC) is often added to the transmitted data[8]. The *modulation and coding schemes* (MCSs), which are combinations of modulation schemes and coding rates, are characterised with two metrics: data rate and signal robustness. Naturally, there is an inherent trade-off between the two; in fact, bandwidth efficiency and signal robustness are inversely proportional. A higher order modulation and coding scheme attains higher data rate but also higher error rate over lower MCS for the same radio channel conditions.

As aforementioned, the earlier generation of mobile systems have not been optimal in terms of efficient resource use because these incorporated a single modulation that achieved prescribed error rate under the worst propagation conditions. Note that the very nature of the wireless channel does not allow the *steady use* of a bandwidth efficient modulation therefore, the earlier wireless systems were dimensioned for conditions experienced at the far edge of the cell. A major step in overcoming such inefficient design was the introduction of the *adaptive modulation and coding* (AMC) technique in the advanced mobile wireless system. AMC takes advantage of favourable radio propagation conditions when these are present and decreases the order of the modulation and coding scheme when the radio link conditions are deteriorated (see Fig. 6.1); that is, it dynamically adapts the modulation scheme and coding rate to the radio channel state. Hence, AMC facilitates dynamic operation near the optimal point over a time-varying channel, rather than optimally for a static channel model. Adapting to the propagation conditions can increase the average throughput, reduce the re-

[8] When the propagation conditions are ideal in theory and excellent in practice, the use of FEC can be avoided.

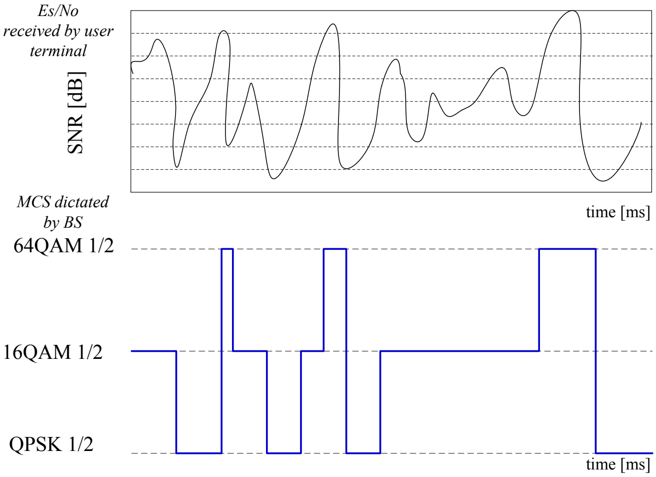


Figure 6.1: Adaptive modulation and coding mode

quired transmission power by taking advantage of good signal quality (to send at higher data rates or lower power), or reduce the probability of bit error by using a robust modulation and lower coding rate. The main goal is to maximize the average spectral efficiency while maintaining a given average or instantaneous bit error probability.

Typically, several different *burst profiles* (or *modes*, i.e., MCSs) are defined and the operation regions for each of them are quantitatively defined. Different burst profiles are used depending on the experienced radio propagation conditions. These conditions are assessed using suitable metrics as reliable link quality indicators. In practice, the adaptive modulation and coding algorithms often implement *signal to noise ratio* (SNR) or *signal to interference ratio* (SIR) depending on the predominant wireless conditions. The system dynamically tracks the time-varying radio channel and depending on the measured level of the signal quality indicator determines the burst profile. Fig. 6.1 illustrates the principal premise of the adaptive modulation and coding technique. For simplicity three modulation schemes and one coding rate are considered. If the signal is equal or above the SNR threshold for the given MCS (say 16-QAM) then this scheme is used (i.e., 16-QAM). If however, the mesured SNR is below the aforementioned threshold than a lower-order MCS (QPSK in the example) is applied.

To complement our discussion on AMC it is important to note that there are a few basic requirements for its practical implementation [43]: (i) a (fast) feedback between the transmitter and receiver, (ii) estimation and feedback of channel quality faster than the changes observed in the radio link; (iii) transmitter that can quickly and often change its rate and power. In addition to the aforelisted requirements, Goldsmith [43] points out that the quality of fixed-rate applications with hard delay constraints (as voice or video for example) may be considerably compromised. This issue is a focus of research in the chapter following.

In practice, the adaptive modulation and coding technique was implemented in the 3G networks, such as *evolution-data optimised* (EV-DO) and *high-speed data access* (HSDA). Based on the favourable results and practical experience, its application has been extended beyond these technologies and AMC has been incorporated in the IMT-Advanced systems, namely LTE and WiMAX.

### HARQ

For achieving enhanced and reliable transmission of packets *automatic repeat request* (ARQ) technique is commonly adopted at the link layer. Naturally, the use of ARQ generally depends on the experienced radio channel conditions: if these are far from ideal, ARQ can greatly improve system performance; otherwise (when link quality is good), ARQ is usually not used because the retransmissions cause a reduction in the overall throughput.

At the receiver, the error detection code of each data packet is used to determine if one or more bits of the packet are corrupted[9]. In general, if these cannot be corrected the receiver discards the packet and informs the transmitter via a feedback channel that the packet must be retransmitted. When *hybrid*-ARQ[10] is used, the packet instead of being discarded is saved and different strategies are posteriori used for its correct reception.

[9] The packtets contain *cyclic error correction* (CRC).

[10] The HARQ is part of the IEEE 802.16 familiy of standards.

When *chase combining* technique is implemented, the corrupted packet is combined with the retransmitted one. Alternativelly, rather then retransmitting the original packet, some additional coded bits are send by the transmitter to provide a stronger error correction probability. The later technique is called *incremental redundancy.* HARQ supports improved link performance compared to the traditional ARQ technique but at the cost of higher implementation complexity.

### MIMO

Multiple-antenna radio systems also referred to as *multiple in multiple out* (MIMO) systems are another enhanced technique that can substantially improve link reliability and spectral efficiency of the wireless communications due to which MIMO is adopted by IEEE 802.16e standard as well as by other broadband wireless standards, such as Wideband Code Division Multiple Access (WCDMA), LTE, IEEE 802.11 (WiFi), and IEEE 802.20. In effect, there are two different ways through which MIMO brings these improvements. Specifically, MIMO is not aimed at mitigating the adverse propagation conditions but rather takes advantage of the rich multipath environment (also referred to as diversity gain) to achieve signal robustness in terms of bit error rate. Signal diversity can be obtained through multiple transmit antennas[11]. Another MIMO technique is the spatial multiplexing, which uses independent signalling paths to send independent data in parallel. As with the HARQ technique, the attained spectral efficiency gains and reduced (inter-symbol and user) interference are at the cost of implementation complexity (e.g., hardware complexity in deploying multiple antennas, especially in handheld devices).

## 6.3 Mobile WiMAX technology

In the Chapter 7 following we focus our research on the point-to-multipoint WiMAX configuration as for this operational mode the base station cen-

[11] Space-time coding on the other hand, addresses the mutual interference generated of simultaneously transmitting antennas.

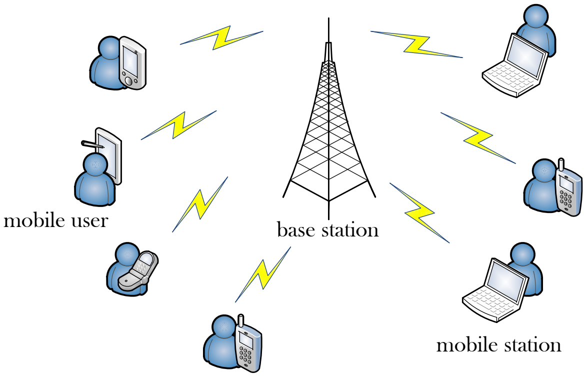


Figure 6.2: Mobile WiMAX cell organization

trally manages the allocation of resources to incoming call requests and thus, facilitates quality of service support. In particulr, an IEEE 802.16e cell consists of a base station that serves several mobile stations as illustrated in Fig. 6.2. Before transmission to the base station can start, the mobile station must request admission of the new connection (hereafter denoted *call* for short). If accepted, the base station is then responsible for meeting the QoS requirements of the call. Note that the scheduling discipline as well as the particular admission control algorithm, i.e., the responsible modules for providing QoS support, are not defined in the standard [2] but left open to vendor specific solutions. Time is partitioned into frames of (usually fixed) dimension, which are divided into downlink and uplink subframes. A transmit transition gap (TTG) and receive transition gap (RTG) separate the downlink and uplink subframes in the time-division duplex (TDD) mode of the Wireless MAN OFDMA (the one portrayed in Fig. 6.3).

The frame is two dimensional (see Fig. 6.3): time (on the axis) in units of OFDMA symbols and frequency (on the ordinate) in units of

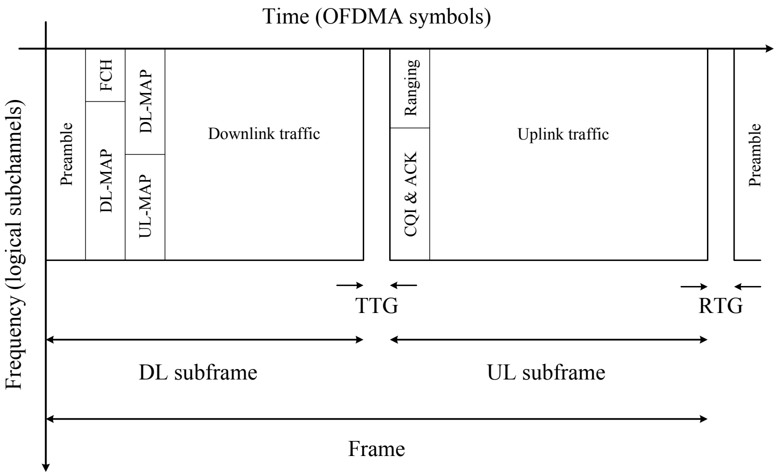


Figure 6.3: TDD frame structure of mobile WiMAX

logical subchannels (see Section 6.2). The number of symbols and subchannels per frame depends on the frame duration, channel bandwidth, direction, and subcarrier permutation among others. The mandatory subcarrier permutations—the mapping of logical subchannels onto physical subcarriers—defined by the standard are *partial usage of subchannels* (PUSC) and *full usage of subchannels* (FUSC). The basic resource allocation unit in OFDMA is called *slot* and consists by at least one symbol and at least one subchannel. The number of slots per frame is fixed. As mentioned above, the slot structure depends on the particular subcarrier modulation used. Assuming for instance PUSC, a *downlink* (DL), from base station to mobile station, slot is composed by two symbols and one subchannel. The *uplink* (UL), from mobile station to base station, slot is determined by three symbols by one subchannel. Whereas the structure of the slot depends on the subcarrier permutation, the capacity of each slot (the amount of data that can be transmitted, or the bit per slot rate) depends on the modulation and coding scheme, see Table 6.1. The IEEE 802.16e standard specifies several burst profiles (i.e., MCSs) to be used by the mobile station and base station to adapt to the radio propagation conditions. The base station advertises on a regular basis the set of available MCSs through the *downlink channel descriptor* (DCD) and *uplink channel descriptor* (UCD) control messages. Moreover, the base station decides on the burst profile per connection and per frame; that is, the modulation and coding scheme for each connection are determined on a frame basis. Mobile stations send the measured signal quality to the base station, which *channel quality information* (CQI) can be feeded back on the *channel quality information channel* (CQICH). The base station on the other side measures the quality of the received signal as well and decides on the modulation scheme and coding rate to be used for downlink transmission in addition to the MCS for the uplink; the latter decision is based on the received at the base station signal quality data measered at the mobile station.

The downlink frame starts with a preamble for synchronisation and channel estimation purposes, which consists of a known sequence of modulated pilot frequencies. The *frame control header* (FCH) carries system control information including the length of the DL-MAP. The base station specifies the frame allocation (the allocation of data regions[12] to users) in the *downlink map* (DL-MAP) and *uplink map* (UL-MAP) messages, which are advertised at the beginning of the downlink subframe. For the DL-MAP and UL-MAP the most robust MCS is employed to ensure the correct decoding of the messages from all mobile stations independent of their location and propagation conditions.

The mobile WiMAX specification allows virtually all available spectrum width to be used. In particular, the channel width can be in the range from 1.25 MHz to 28 MHz. The channel is divided into equally spaced subcarriers. In the case of a 10 MHz channel, which we consider in the numerical evaluation part of Chapter 7, there are 1024 subcarriers the majority of which are used for data transmission, whereas the remainder are used for control (monitoring the quality of the channel, providing safety zones, or a reference frequency). The default values recommended by the mobile WiMAX Forum include a 10 MHz channel, TDD mode, 5-milisecond-frame,

[12] *Data regions* are a two-dimensional allocation of a group of contiguous subchannels, in a group of contiguous OFDMA symbols [2].

Table 6.1: Slot capacity for various MCSs [92]

| MCS | Bits per symbol | Coding rate | DL bytes per slot | UL bytes per slot |
|---|---|---|---|---|
| QPSK 1/8 | 2 | 0.125 | 1.5 | 1.5 |
| QPSK 1/4 | 2 | 0.25 | 3.0 | 3.0 |
| QPSK 1/2 | 2 | 0.50 | 6.0 | 6.0 |
| QPSK 3/4 | 2 | 0.75 | 9.0 | 9.0 |
| QAM-16 1/2 | 4 | 0.50 | 12.0 | 12.0 |
| QAM-16 2/3 | 4 | 0.67 | 16.0 | 16.0 |
| QAM-16 3/4 | 4 | 0.75 | 18.0 | 16.0 |
| QAM-64 1/2 | 6 | 0.60 | 18.0 | 16.0 |
| QAM-64 2/3 | 6 | 0.67 | 24.0 | 16.0 |
| QAM-64 3/4 | 6 | 0.75 | 27.0 | N/A |
| QAM-64 5/6 | 6 | 0.83 | 30.0 | N/A |

PUSC subchannelisation and DL : UL ratio 2 : 1.

As mentioned above, the IEEE 802.16e MAC protocol is connection-orineted. Furthermore, the connections are uni-directional, either from the base station to the mobile station (forward path or *downlink*), or from the mobile station to the base station (backward path or *uplink*). A mobile station can establish more than one simultaneous calls with the base station.

The MAC protocol explicitly supports QoS by defining five QoS data delivery services [2], namely *unsolicited grant service* (UGS), *extended real time polling service* (ertPS), *real time polling service* (rtPS), *non-real time polling service* (nrtPS), and *best effort* (BE). These MAC classes are used to support a specific class of applications. The UGS class, which is considered in Chapter 7, has been designed for real time service flows that transport fixed-size data packets on a periodic basis, that is, with very stringent delay and jitter requirements, such as speech (*voice over IP*, VoIP in specific) [2]. Fixed-size grants on a periodic basis are allocated to UGS services to meet the requirements of the real time traffic. These periodic allocations depend on the particular UGS traffic flow metrics: minimum reserved traffic rate, tolerated jitter, and maximum latency (see [2]) and are negotiated during the initiation of the call. The other mandatory QoS parameters of the UGS classs are: uplink grant scheduling type, *service data unit* (SDU) size, request/transmission policy, and unsolicited grant interval.

# 7 Effect of AMC on system-level performance and admission control design in mobile WiMAX networks

## 7.1 Introduction

The conventional mobile cellular networks, which were a focus of research in the first part of the thesis, are voice-oriented and importantly, are devised for the worst-case signal conditions scenario. The allocation of resources (traditionally called *channels*) to call requests in the these networks hence, is deterministic. In contrast to the fixed resource demand in the conventional land mobile systems, the resource demand in the present and future (specifically, the IMT-Advanced) networks is not deterministic but inherently dynamic. The latter is due to new mechanisms introduced in the wireless systems to improve their performance. In particular, in order to support larger service diversity and applications with significantly higher compared with the plain voice service rate demands, the capacity of the wireless networks was expanded through spectrum-efficient techniques. The *adaptive modulation and coding* (AMC, sketched in Chapter 6) technique for instance, has been widely adopted by the advanced wireless communication standards because it attains more efficient use of the scarce radio resources. However, from the perspective of admission control, the adaptive modulation and coding technique introduces a fundamental difference between the earlier generation wireless networks and the mobile broadband systems namely, the deterministic resource use of a constant bit rate service is translated into a dynamic resource demand that depends on the experienced radio channel conditions as explained later in this chapter.

In Part II of the thesis we centre our investigations on studying the effect of the dynamic resource demand on the system model of and importantly, admission control design in mobile cellular networks with broadband access, which research question has not yet been investigated in depth in the open literature. Our approach is analytical as it allows us to isolate the effect of AMC from the rest of the parameters and clearly define its importance for different radio link, mobility and traffic conditions.

In the remainder of this introductiory section the relevant literature is overviewed, the motivation and goals of the research reported in the sections following are further explained, and the main contributions are identified (see Section 7.1.1). Then, the background for zone-based cell modelling is provided, and the model and input metrics of the analysed system are defined (see Section 7.2). Next, the two general approaches for admission control design adopted in the WiMAX literature are modelled (see Section 7.3). Later, the performance measures for mobile WiMAX system performance evaluation are established (see Section 7.4). The numerical results derived from this study are plotted and analysed in Section 7.5. Finally, this chapter is concluded with a summary of the reported research and main findings (see Section 7.6).

### 7.1.1 Background

In the previous chapter we briefly overviewed some of the recent technological advances in mobile communications bearing in mind the mobile WiMAX technology because WiMAX is representative for the foremost broadband mobile standards. We paid a special attention to the adaptive modulation and coding technique because the available system capacity and resource demand, which are main metrics in admission control, depend on the AMC-defined bits-per-slot rate[1]. Consequently, compared with the classical model of a mobile telephone cellular network considered in Part I, the mobile WiMAX system model differs significantly. In particular, recall that the classical model typically used by the central body of handoff prioritisation algorithms (see [5, 31, 42, 88, 104]) is characterised with deterministic resource allocation, as the conventional cellular systems are devised for conditions experienced at the far edge of the cell and do not adapt to the fluctuating radio channel conditions. As a result, the amount of resources that a call needs remains fixed for its duration. With respect to the traditional system model, the AMC technique converts the system capacity from a fixed parameter to a random variable. The time varying capacity when measured in terms of achievable bit rate depends on user's location in the cell and experienced radio link conditions. When the capacity is determined in terms of the maximum number of communication sessions it depends also on the service type (streaming or elastic) and quality of service requirements. In an effort to quantify mobile WiMAX system capacity So-In *et. al* [92] derive guidelines for capacity computations. The authors show with various examplifying scenarios that among other factors such as service type, QoS requirements, designer specific solutions (regarding the scheduling algorithm, mechanisms for overhead reduction, silence suppression, etc.), WiMAX system capacity strongly depends on users' location and experienced signal conditions through the assigned modulation and coding scheme (MCS). The VoIP system capacity for example, for the recommended by WiMAX Forum system configuration [109], is of 42 simultaneous VoIP users when a QPSK 1/8 scheme is applied and 82 VoIP users for the less restrictive QAM-64 5/6 scheme[2] [92]. The proposed theoretical method is useful for understanding the effect of various parameters, inclu-

[1] Recall from Chapter 6 that the slot is the main resource unit in WiMAX and that its capacity (or bit rate) depends on the modulation and coding scheme; that is, slot's capacity is dynamically determined by the AMC technique.

[2] It is assumed that all mobile terminals are assigned the same modulation and coding scheme [92]; that is, experience the same radio link conditions. Other examples that clearly show the difference in number of served connections when different MCSs are used can be found in the same publication [92].

ding that of the burst profile. However, it is developed assuming a single modulation and coding scheme and it does not study nor provide insights into the effect of the adaptive modulation and coding technique on system-level performance. Furhter, So-In *et. al* discuss the implications of the used modulation and coding scheme on scheduler resource allocation [94], and suggest that WiMAX system models and practical solutions should consider the AMC induced time-varying bits-per-slot rate [91, 93, 94].

### Literature

The WiMAX admission control studies that prioritise active calls in front of new call requests in order to support the continuity of the former can be classified into two major groups based on the assumptions made about the resource environmnent. The first group of admission control proposals assumes fixed resource allocation (that is, ignore the variablity of the resource demand), whereas the proposals from the second group model the time varying radio link conditions and incorporate them in admission control design. The first group of admission control proposals is very similar to the existing solutions for traditional cellular networks because the same deterministic resource conditions and as a result system model are assumed except for the fact that instead of considering only constant bit rate applications (streaming traffic or voice calls), other service classes (generally denominated elastic traffic) are incorporated. As most of these solutions do not differ conceptually from the traditional approaches (see Chapter 3 and the cited compendia there) we limit our discussion here only to those from the second group. Note, however, that later on in this chapter we compare and contrast the two general classes of admission control approaches found in the WiMAX literature.

The studies on admission control in mobile WiMAX systems that model the variations in slot capacity and consequently consider them for divising admission control algorithms are scarce. Kwon *et. al* [62, 63] are the first to contribute in the field by modelling a mobile WiMAX system with AMC. The authors relax the fixed channel capacity assumption and propose a modified guard channel scheme that allows both types of active calls—modulation-changed and handoff—to use the guard capacity. Their numerical evaluations show that such a strategy in general decreases the dropping probabilities of active calls.

In a series of publications Tarhini and Chahed [17, 98–103] evaluate WiMAX system-level performance as well. In [99] the authors present an AMC-aware QoS proposal that gives the highest priority to streaming flows: according to the authors the proposed QoS paradigm makes OFDMA-based IEEE 802.16 capable of offering a sustainable rate to streaming applications throughout the whole coverage area. Specifically, different amounts of resource units are allocated to streaming calls belonging to the different MCS-defined regions to achieve the same (for streaming calls) constant bit rate within the cell area. The left-over capacity is assigned to elastic traffic calls, thus, the proposed AC policy can be considered a complete sharing with service priority assignment scheme. It is important to note that the research in [99, 100] is devoted to a fixed scenario: that is, the users are static and the experienced radio link conditions are assumed to be constant. The model and concept described in [99] are extended later to a single cell mobility case: intra-cell but no inter-cell mobility is assumed. As a sequel of these investigations Tarhini and Chahed porpose a density-based admission control scheme that prioirtises streaming ongoing calls over elastic traffic [101,103] for WiMAX systems with adaptive modulation and coding technique. The acceptance criterion of a new call request depends on two parameters: the already accepted and maximum number of concurrent calls in a given MCS-determined region.

Wang and Iversen [107] model as well the implementation of adaptive modulation and coding in WiMAX systems. In particular, the authors model a multi-user, multi-class OFDM-TDMA system with AMC. The adaptive admission control strategy proposed in [107] deals with the outage probability that a call that suffers a change to a more robust modulation and coding scheme cannot be accommodated in the system because of insufficient resources and as a consequence is dropped. It is assumed that each service class shall be guaranteed a certain bit rate. An analytical approach is proposed for determing the time congestion, call congestion, and traffic congestion probabilities. The study concentrates on an isolated cell case, where call drop can be evidenced due to deterioration of the signal quality of the radio link *and* simultaneous resource shortage. Note, however, that inter-cell handoffs are not modelled. It is worth mentioning that the inter-cell mobility (inter-cell handoffs), inherent in mobile cellular systems, is essencial when evaluating cellular performance (see Part I). In fact, not only [107] but the research on WiMAX networks with AMC in general (see [99–101,103]), except for [62,63], is carried out for a single cell scenario with intra-cell mobility but not for a mobile WiMAX system, where inter-cell handoffs in addition to intra-cell handoffs[3] can take place. In [62, 63], where inter-cell mobility is considered, only dropping probabilities (due to both inter-cell and intra-cell handoffs) are analytically derived; the forced call termination probability is not analysed.

### Fairness

In conventional mobile cellular systems the capacity of each channel (resource unit) is constant and the capacity allocation for each call is fixed as explained earlier. Furthermore, the system is commonly modelled mono-service and the network is assumed to be homogeneous (see Chapter 2). Hence, calls are differentiated based on their origination class (*new* or *handoff*), and admission control algorithms are typically fair, as all handoff (new) calls are assigned the same amount of resource units (usually one channel), attain the same bit rate, and suffer the same forced call termination and new call blocking probabilities throughout the service area. In mobile cellular networks with broadband access, such as those based on the WiMAX technology, the resource allocation is non-deterministic even for applications with constant bit rate (*cbr*) requirements[4] as the amount of needed resource units depends on the user location and signal quality of the radio link. Therefore, in mobile WiMAX networks (packet-level and system-level) unfairness, to be defined shortly, between calls that experience good and calls that suffer unfavourable radio channel conditions is often observed.

The unfairness problem has been primarily studied on packet level

[3] The switching between modulation and coding schemes is commonly modelled as *intra-cell handoff* as explained later in the chapter.

[4] At the MAC layer, application services with constant bit rate requirements are mapped to the UGS class [2], as described in the previous chapter.

[6, 7, 35, 36, 94]. *Packet-level fairness* is a measure of the capability of the scheduler to achieve temporal or throughput fairness [36]. Ahmed *et. al* [6,7] study short-term throughput fairness among different users in fixed broadband wireless access networks with power control and adaptive modulation and coding technique. *Short-term throughput* fairness is measured through the throughput attained by each user on a frame basis [6, 7]. Fallah and Alnuweiri [36] introduce the notion of (long-term) throughput and temporal fairness and examine those in a multirate IEEE 802.11e wireless local area network (WLAN). Fallah and Alnuweiri denote *temporal fairness* those resource allocations in which each user is assigned the same (weighted) amount of service time regardless of their transmission rate [36]. In contrast, *throughput fairness* is provided when all users obtain the same throughput in a given time interval [36]. Fallah *et. al* [35] study fair scheduling in IEEE 802.16 for real-time multimedia support by considering the explained fairness concepts and proposing fair scheduler solutions. So-In *et. al* [94] adopt the notion of temporal and throughput fairness and adapt them to the context of mobile WiMAX networks (*slot* and *byte* fairness respectively). The authors propose a generalised weighted fairness scheme for scheduler incorporation. Such implementation is done for instance in [93]. After defining the notion of system-level fairness it will become evident that the overviewed proposals [6,7,35,36,94] can solve the fairness problem on a packet level but not on a system level.

Unlike the packet-level (un)fairness, the system-level (un)fairness in mobile WiMAX has not been studied before. *System-level fairness* though, is a fundamental metric as it measures the capability of the system to provide uniform or below agreed levels new call blocking and forced call termination probabilities within the service area; that is, it is a measure of the system-level QoS guarnatees that the network can provide throughout the coverage area. It is relevant to examine system-level fairness because it can be perceived by the mobile user: it would be unacceptable, from the mobile user perspective, to experience dissimilar and (or) unacceptable blocking (dropping) probabilities in certain regions of the system. Mobile user's expectations and requirements in terms of both service diversity and quality have evolved and nowadays, the mobile user requests equivalent or better QoS that the one offered by the wired networks. Further, the mobile user expects to receive QoS independent of where in the network the communication takes place. The very notion of ***mobility*** suggests that the mobile user should not be concerned with their location or direction of movement while engaged in a communication session. Mobile cellular networks, especially the next generation, hence, shall be designed to meet the user expectations for ubiquous service and quality levels.

Despite the central importance of providing system-level fairness in mobile cellular networks, this problem has neither been examined nor tackled in the overviewed works on mobile WiMAX [17,62,63,98–103,107]. In Kwon *et. al* blocking (dropping) probabilities of calls belonging to different cell regions are not equalised, which is also valid for the works of Tarhini and Chahed [17,99–103]. Moreover, it is apparent that the proposed by Tarhini and Chahed [101, 103] density-based admission control is biased towards calls that experience better radio link conditions and we argue that this yields higher disparities between blocking (dropping) probabilities of calls from different cell zones. The adaptive algorithm of Wang and Iversen [107] aims at maintaining the outage probability (probability that an active call will be dropped due to insufficient resources) below a given level: a new connection is accepted with some probability provided that if admitted the system outage probability will be maintained. The blocking probability in [107] is a random variable that depends on the state of the system and the *bandwidth requirement* of the new connection. Thus, the new call blocking probabilities of different service classes are inherently dissimilar.

It is noteworthy that although not studied in WiMAX networks, in the past, the relevance of the system-level fairness problem was not left unnoticed but has attracted researchers' attention and effort. Naturally, it has been investigated in conditions that lead to new call blocking and handoff call dropping unfairness. It has been examined, for instance, in GSM networks with half- and full-rate connections [53] and in multimedia mobile cellular networks with fixed resource allocation (see [34] and references therein). These studies elaborate on the importance of providing uniform or equal (in terms of blocking and dropping probabilities levels) system fairness. Furthermore, Ivanovich [53], Epstein and Schwartz [34] analyse and design admission control algorithms from the perspective of fairness and efficiency as critical measures for the studied mobile cellular networks. More recent studies that divise admission control policies for heterogenous networks [52, 60] address *inter-cell* system-level unfairness—high disparity between blocking probabilities in hot-spot and moderately loaded cells—and *inter-service* system-level unfairness—discrepancy between blocking probabilities of voice and data services.

As it shall be explained shortly after, we assume homogeneous network and consider the UGS WiMAX delivery class, for which conditions inter-cell and inter-service unfairness are not manifested but unfairness due to different MCS is observed.

### Contributions

Our *goal* is to study the effect of the adaptive modulation and coding technique on system-level performance and admission control design for constant bit rate service as noted previously. This research was motivated by the fact that despite investigations on WiMAX system-level performance, several important questions have remained unanswered:

- Is it a significant hypothesis to ignore that the resource demand induced by the time-varying wireless link conditions is *dynamic* and assume that the resource environment is *fixed*?
- How AMC impacts cellular performance in terms of blocking and forced call termination probabilities? Under what conditions?
- What are the implications of the non-deterministic resource environment on admission control design?

The first question was *motivated* by the two general approaches found in the WiMAX literature, namely works that make simplifying assumptions about the resource environment and consider it fixed and studies that relax this assumption and model its dynamics as explained earlier. To the best of our knowledge, there is no study that assesses the impact of this simplifying hypothesis on system performance. Furthermore, the open literature on mobile WiMAX lacks investigations that specify the radio link, mobility, and traffic conditions for which AMC has noticable effect on system-level performance, which motivated the second question. Next, it is of much practical interest to determine how admission control design depends on the adaptive modulation and coding technique.

In the light of the preceding discussions our main *contributions* are as follows:

- We specify the effect of the adaptive modulation and coding technique on system performance: that is, explain how it is manifested and what phenomena are observed.
- We identify conditions for which the dynamic resource demand due to the time-varying wireless link significantly alters system performance.
- We elaborate on system-level (un)fairness as a consequence of the dynamic resource demand, a question not addressed previously in the WiMAX literature.

Based on our investigations we derive general guidelines that can be useful in devising admission control schemes.

We undertook an analytical instead of a simulation study of the questions set above because the mathematical analysis facilitates modelling of a particular feature, simplifying the rest of conditions, which eases the focus on and evaluation of just the parameter of interest. This is especially valid for the WiMAX technology that incorporates many advanced techniques, which can prevent the effect that the dynamic demand has on cellular performance from being clearly observed. We have considered carrying out simulations to complement our work on a second stage; specifically, to evaluate AMC under the combined effect of several WiMAX technology specifics (see [106]).

## 7.2 Analytical framework

In the literature, one of the most often used analytical models that captures the different propagation conditions experienced within a cell is the zone-based model. We shall assume this same cell model in the sections to follow, thereby we provide an overview of the relevant literature first and then we proceed with specifying the particular model our investigations are based on.

### 7.2.1 Zone-based cell model

Many publications that address different aspects of the wireless communication systems model the area covered by a base station by splitting it into concentric zones, see Fig. 7.1. In principle, it is assumed that each zone of the cell can give different support from the rest of the zones. Jiang and Rappaport [56] for example propose a concentric division of the cell into circles in order to model a system with a hybrid channel allocation (HCA). The authors divide the cell into two concentric zones: an outer one with fixed resources (i.e., FCA) and an inner one where the base station resources are allocated in a dynamic fashion (i.e., DCA) depending on the traffic load [56]. Su *et. al* propose the same concept of cell modelling [97] but for evaluating a CDMA cellular network performance with soft handoffs. A cell in the considered network is approximated by a circle and cell's area is divided into two regions: normal (modelled by a circle) and handoff (modelled by a ring). The authors evaluate the impact of the soft handoff on cellular performance [97]. Their model was later extended by Zhang and Lea [114] to study the mobilty induced interference in CDMA cellular network with soft handovers. Elayoubi *et. al* [33] propose the same abstraction—decomposition of a cell into concentric rings—to model the interference levels encountered by users in different regions of a cell belonging to a CDMA network.

The cell modelling concept fundamental for the described works has been also implemented in studying the effect of link adaptation techniques in cellular networks. Specifically, Khan and Zeghlache [59] refer to the model proposed by Jiang and Rappaport [56] and apply it for evaluating the improvement in system capacity of a GSM network that serves half-rate voice in addition to full-rate voice codec calls. Explicit mathematical formulae for the research problem considered by Khan and Zeghlache [59] are derived by Cruz-Perez *et al.* [26]. Later, Cruz-Perez *et al.* systematically apply the model in their posterior works to evaluate other problems related to mobile cellular networks including residence time distribution in a cell zone [28, 80], link adaptation [29], reuse partitioning [68], performance of CDMA systems [81], etc. Kwon *et. al* [62, 63], as well as Tarhini and Chahed (see for example [100]), whose works we discussed earlier in this chapter, propose the same abstruct idea to model a WiMAX system with adaptive modulation and coding technique. Tarhini and Chahed in particular, extend the concept of modelling the radio propagation conditions in a cell through cell division into concetric regions suggested in [33] to a WiMAX network with AMC.

Although the previously described research works focus on different aspects of mobile cellular systems, common for the majority[5] of them is that the same mathematical abstruction is adopted to model the area covered by a base station, that is, a cell. In our mathematical analysis we implement the basic principle of spliting the cell into concentric regions in order to model the varying propagation conditions, and consequently the use of different modulation and coding schemes that depend on the wireless channel signal quality.

### 7.2.2 Assumptions

The propagation conditions over the radio medium depend on several factors but in wireless communication studies these are confined to three major propagation effects [43,87]: (i) *path loss*: the reduction in the average power strenght of the transmitted radio signal as it propagates in space, expressed as a power law of the distance between transmitter and receiver[6], (ii) *shadow* (*log-normal fading*): long-term variation around the average power

[5] The work of Wang and Iversen [107] discussed in the previous section diverges form the widespread cell-zone modelling approach. In particular, the authors adopt the idea proposed in [65]: assume a Rayleigh fading channel for which the SNR has a known probability density function, devide the entire SNR range into non-overlaping successive intervals for which different modulation and coding schemes are defined so that a prescribed packet error rate can be maintained. Finally, the number of bits that can be transmitted over a given subchannel with some SNR in one time slot (in essence OFDM slot capacity) is defined as a random variable.

[6] $1/(d)^n$, where $d$ is the distance, and $n$ is the path loss exponent, typically ranging between 2 (free space) and 4 (for urban cellular networks) [85].

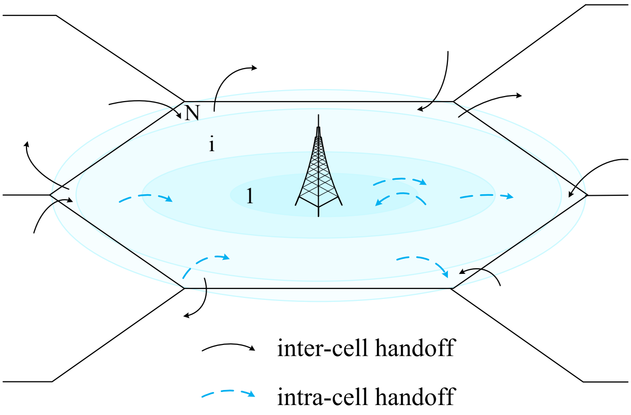


Figure 7.1: Zone-based model of a cell

over distances of the order of the lenght of the obstacle, and (iii) *fast fading* (*multipath fading*): rapid signal variations over small distances (of the order of the wavelenght) manifested mainly when the users are moving.

The propagation model most often used to define the propagation conditions experienced in different parts of a cell is the free space model [26, 59, 63, 99]. The received power for the free space propagation model is given by [87]:

$$P_R = P_T G_T G_R \left( \frac{\lambda}{4\pi d} \right)^2,$$

The assumption that the propagation conditions depend mainly on the distance between the transmitter and receiver lead to a simple model of the wireless medium that is otherwise difficult to characterise. We make the assumption of a free space propagation.

Other assumptions that we make about the system model are as follows:

- A point-to-multipoint WiMAX configuration is considered as for this operational mode the base station centrally manages the allocation of resources to incoming call requests, which facilitates QoS support.

- It is assumed that stationary, homogeneous conditions prevail in the considered mobile WiMAX cellular system. This means that the mobile stations are uniformly distributed throughout the network coverage area and each cell observes the same traffic and mobility patterns[7]. As a consequence of the assumption that all cells are identical we can model the performance of a single cell.

- We do not model the use of a MIMO antenna system; instead, we assume that omnidirectional antennas are placed at the centre of the BS service area.

- The shape of the cell is approximated by a circle. We assume a free-space propagation model as pointed out previously; that is, the signal quality depends only on the distance between the base station and mobile station as clarified above. As a result, the cell can be split up into concentric circle and rings to model the propagation conditions experienced by users in different parts of the cell (see Fig.7.1). The signal conditions are assumed to be the same inside every ring.

- The same modulation and coding scheme is used in *downlink* (DL), from base station to a mobile station, and *uplink* (UL), from a mobile to base station, as commanded by the latter. This is a consequence of the assumption that the wireless channel conditions (i.e., signal to noise ratio) are the same inside every ring.

As a mobile technology WiMAX keeps the main characteristics described in Chapter 2, Section 2.1, i.e., cellular layout, overlapping areas, handoff mechanism, mobile users that request service intermittently. Moreover, mobile WiMAX is specified by:

- base station radio resource assignment

  One of the key advantages of the OFDMA air interface is the possibility of enabling *frequency reuse of one*; that is, the same frequency

[7] Recall that homogeneous cellular network was just the case we discussed in Part I of the thesis.

can be used in all neighbouring cells and sectors[8].

- call resource demand

  It is worth recalling that we consider a mobile WiMAX that adopts the adaptive modulation and coding technique. The latter allows the system to adjust the modulation scheme and coding rate dynamically in response to the frequently changing radio channel conditions: more robust MCS is used when radio channel conditions are poor, whereas more spectral-efficient MCS is used when the conditions are favourable. According to IEEE 802.16, the BS decides on the burst profile (that is, the MCS) on the downlink and on the uplink based on the current signal to noise ratio (SNR). The decision is made on a per-user and a per-frame basis [2]. Therefore, the call resource allocation is *non-deterministic* even for the considered in our study applications with constant bit rate requirements such as voice service.

### 7.2.3 System model

#### Input parameters

The system model that we define below is described by the following parameters:

- $C$ – total number of system resources (in slots)
- $s_i$ – slot capacity in zone $i$
- $g$ – number of guard units
- $C_g$ – number of guard slots

  The guard resources are calculated in terms of slots and are given by:

$$C_g = g \cdot s_N,$$

[8] High frequency reuse patterns cause interference limitations to the systems therefore, other feasible implementations are also considered such as 1 x 3 x 3. The latter corresponds to cell sectorisation in three, each sector having a different frequency allocation, while the reuse factor between cells is 1 again.

where $s_N$ is the slot capacity in the outermost zone (zone $N$)

- $n_i$ – number of active users (ongoing calls) in zone $i$
- $1/\mu$ – mean of the unencumbered call duration time
- $1/\eta$ – mean of the cell residence time
- $1/\eta_i$ – mean of the residence time in zone $i$
- $\lambda_n$ – mean of the total new call arrival rate
- $\lambda_n^i$ – mean of the new call arrival rate in zone $i$

  We assume that the mobile users are uniformly distributed within the service area therefore, the mean of the new call arrival rate in each zone can be determined through the total new call arrival rate and the zone area. It is given by:

$$\lambda_n^i = p_i \lambda_n, \tag{7.1}$$

  where $p_i$ is the area of zone $i$ expressed as a fraction of the total cell area.

- $\lambda_h^{inter} = \lambda_h$ – mean of the inter-cell handoff call arrival rate

  It is determined by applying an iterative procedure as explain later in this section.

- $\alpha$ – mobility factor or handoff rate

  In analytical modelling the handoff rate is usually approximated by the cell residence time and call holding time through the following expression:

$$\alpha = \frac{\eta}{\mu} \tag{7.2}$$

- $\lambda$ – mean of the aggregate arrival rate given by:

$$\lambda = \lambda_n + \lambda_h \tag{7.3}$$

- $q_{i+1,i}$ – probability that a mobile user in an outer zone $i+1$ will move to a neighbour zone $i$ that is closer to the base station

We make a common assumption that the new call arrival rate, inter-cell handoff rate, cell and zone residence times, as well as call holding time follow a *negative exponential distribution.* As explained in Chapter 2, the memoryless hypothesis greatly facilitates the analytical work and can be used for exploring general trends.

### Description

In our study, we consider only the unsolicited grant service class, as for UGS we expect to see the effect of AMC manifested unambiguously. Recall that the UGS class has the most severe service requirements. It is specified by the maximum sustained traffic rate, maximum latency tolerance and jitter among others [2]. The rest of the WiMAX scheduling classes have less stringent delay requirements, which gives more freedom to admission control and scheduler designers. The temporal resource insufficiency can be addressed by putting the call in a queue within the limits set by the maximum latency parameter (of ertPS and rtPS) instead of dropping it immediately, or by renegotiating the application's QoS parameters. Such strategies, however, would not allow us to assess the effect of AMC on system performance clearly, which is the reason to focus on the class with the most strict requirements (i.e., UGS). In this line of thought, the analysis performed gives the upper bound on system performance.

Because the UGS class has the most limiting packet-level quality of service requirements, we adopt the policy that every call that cannot be allocated the requested number of slots is immediately dropped. This policy is applied in three cases of receiving a call request: (1) a new call (independent of the zone where it is originated), (2) an intra-cell handoff to a zone with lower-order modulation, and (3) an inter-cell handoff.

Also motivated by the fact that the UGS delivery class has very stringent packet-level service requirements contrary to the rest of the WiMAX classes, it seems reasonable to expect that operators will use different concepts of admission control strategies—complete sharing, complete partitioning and combinations of these—for the different classes. For a mix of complete partitioning and complete sharing strategies, system capacity is divided into two pools (complete partitioning): one explicitly for the UGS class [62, 63] and the other one for the ertPS, rtPS, nrtPS and BE WiMAX classes. The latter pool can be managed by applying complete sharing algorithms that are pertinent to the traffic characteristics of these services (ertPS, rtPS, nrtPS). When the resources from the pool explicitly dedicated to the UGS class are not occupied, they can be used by the rest of the WiMAX delivery services. Another strategy proposed in the literature (see [101, 103] for example) is to give the highest priority to the UGS class, so that for overload conditions UGS calls can occupy the whole system, instead of partitioning the capacity of the system into separate pools. When the focus is on UGS performance as it is in our study both of the aforementioned cases can be represented by the system model discussed below.

Specifically, we model a single-cell in a mobile WiMAX system; that is, with inter-cell and intra-cell mobility. Calls belong to one of two classes: new or active (ongoing). Furthermore, if a call cannot be allocated the requested resources, the call is dropped as noted earlier. The cell is divided into concentric rings in order to model the time-varying and location-dependent propagation conditions within a cell as explained in the preceding section. The system is determined by the parameters described above. The defined system abstruction, allows us to model the dynamic change in call resource usage of an active call inside the cell due to variations in SNR by *intra-cell handoffs* that take place between the borders of adjacent zones of the cell. *Inter-cell handoff requests* model active calls arriving from neighbouring cells (see Fig.7.1). The slot capacity $s_i$ in every zone (the bits-per-slot rate) is defined by the modulation and coding scheme. Naturally, the described analytical model is a simplified version of the real system yet it captures the dynamic resource demand, which is central to our study.

### 7.2.4 Analysis

The system state is described by the number of ongoing calls $n_i$ in each ring $i$ of the cell, that is, by the vector $\vec{n} \equiv (n_1, n_2, ..., n_i, ..., n_N)$, where $i$ is the zone number. The zones are numbered in ascending order starting from the innermost one: zone *1* is the ring immediately surrounding the antenna, zone $N$ is the outermost ring. The system can be modelled by a continuous time Markov chain. The steady state probabilities can be obtained from the set of equations given by:

$$\left| \begin{array}{c} P \cdot Q = 0 \\ \sum P(\vec{n}) = 1 \end{array} \right.$$

The system does not have a product-form solution, therefore the Gauss-Seidel algorithm can be applied for solving it. The state transition probability matrix $Q$ can be determined throughout the transition probabilities between neighbouring states:

$$\begin{array}{l}
p_t(n_1, \ldots, n_i, ..., n_N \rightarrow n_1, \ldots, n_i + 1, \ldots, n_N) = \lambda_n^i \\
p_t(n_1, \ldots, n_i, ..., n_N \rightarrow n_1, \ldots, n_i - 1, \ldots, n_N) = n_i \mu \\
p_t(n_1, \ldots, n_i, ..., n_N \rightarrow n_1, \ldots, n_i + 1, n_{i+1} - 1, \ldots, n_N) = n_{i+1} \cdot q_{i+1,i} \cdot \eta_{i+1} \\
p_t(n_1, \ldots, n_i, ..., n_N \rightarrow n_1, \ldots, n_i - 1, n_{i+1} + 1, \ldots, n_N) = n_i \cdot \eta_i \cdot (1 - q_{i,i-1}) \\
p_t(n_1, \ldots, n_i, ..., n_N \rightarrow n_1, \ldots, n_{i-1} + 1, n_i - 1, \ldots, n_N) = n_i \cdot q_{i,i-1} \cdot \eta_i \\
p_t(n_1, \ldots, n_i, ..., n_N \rightarrow n_1, \ldots, n_{i-1} - 1, n_i + 1, \ldots, n_N) = n_{i-1} \cdot \eta_{i-1} \cdot (1 - q_{i-1,i-2})
\end{array}$$

Note that the transition probabilities are possible provided there are enough resources, a condition that is not explicitly indicated in the above mathematical expressions.

The transition probabilities for the outermost ring $N$ that differ from those specified above are given by:

$$\begin{array}{l}
p_t(n_1, \ldots, ni, \ldots, n_N \rightarrow n_1, \ldots, n_i, \ldots, n_N + 1) = \lambda_n^N + \lambda_h \\
p_t(n_1, \ldots, ni, \ldots, n_N \rightarrow n_1, \ldots, n_i, \ldots, n_N - 1) = n_N(\mu + (1 - q_{N,N-1})\eta_N)
\end{array}$$

Note that an inter-cell handoff can take place only at the outermost ring of a cell. Furthermore, the inter-cell handoff rate is obtained by applying an iterative procedure [59]: after an initial guess of the inter-cell arrival

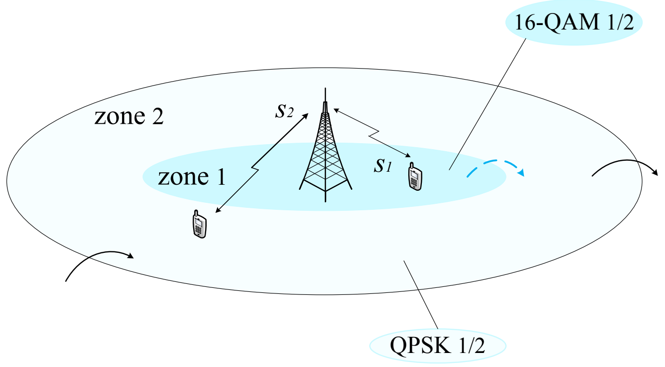


Figure 7.2: A cell with two modulation-and-coding-scheme-defined zones

rate, the equilibrium solution is obtained, and the inter-cell departure rate is calculated from it. The procedure is repeated until the arrival and departure rates converge (a relative error $\epsilon < 10^{-10}$ was used in the numerical evaluations).

It is worth pointing out that the modulation schemes that the IEEE 802.16-2009 standard [2] specifies as mandatory for both links (DL and UL) are quadrature phase shift keying (QPSK) and 16-quadrature amplitude modulation (16-QAM), whereas 64-QAM is optional for the uplink. We assume that the same coding rate is used within the cell so we can focus on a *two-zone* cell model: calls in the inner zone use the 16-QAM modulation scheme, while calls in the outer zone use QPSK (see Fig.7.2). Moreover, we argue that the two-zone cell model is sufficient for assessing the effect of the adaptive modulation and coding rate technique. Indeed, if all the possible coding rates and resulting modulation and coding combinations (i.e., burst profiles) are contemplated, this will lead to a $k$-zone cell model, where $k$ is the number of possible MCSs. We do not expect this $k$-dimensional model to add to the understanding of the problem investigated but to increase the computational effort therefore, we limit our analysis to the two-zone cell

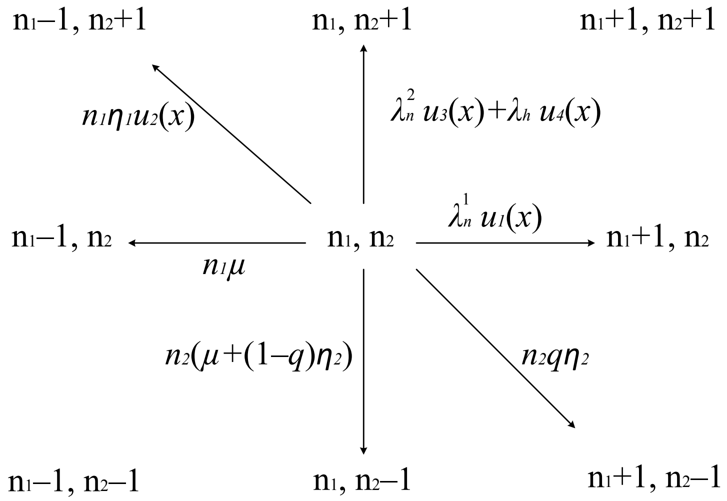


Figure 7.3: State-dependent transition diagram

model. The system state of a cell with two modulation zones is determined by $(n_1,\ n_2)$: number of calls $n_1$ in the inner zone, and number of calls $n_2$ in the outer zone. Its general state-dependent transition rate diagram is shown in Fig. 7.3. The system state $(n_1,\ n_2)$ in equilibrium is given by:

$$\begin{aligned}
&[\lambda_n^1 \cdot u_1(x) + n_2 \cdot (\mu + \eta_2) + n_1 \cdot \mu + n_1 \cdot \eta_1 \cdot u_2(x) +\\
&\lambda_n^2 \cdot u_3(x) + \lambda_h \cdot u_4(x)] \cdot P(n_1, n_2) =\\
&(n_1 + 1) \cdot \mu \cdot P(n_1 + 1, n_2) +\\
&(n_1 + 1) \cdot \eta_1 \cdot u_2(x) \cdot P(n_1 + 1, n_2 - 1) +\\
&[\lambda_n^2 \cdot u_3(x) + \lambda_h \cdot u_4(x)] \cdot P(n_1, n_2 - 1) +\\
&\lambda_n^1 \cdot u_1(x) \cdot P(n_1 - 1, n_2) +\\
&(n_2 + 1) \cdot q \cdot \eta_2 \cdot P(n_1 - 1, n_2 + 1) +\\
&(n_2 + 1) \cdot (\mu + (1 - q) \cdot \eta_2) \cdot P(n_1, n_2 + 1),
\end{aligned} \tag{7.4}$$

where

$$u(x) = \begin{cases} 0, x \leq 0 \\ 1, x > 0 \end{cases}$$

Since there are two zones, $q$ is the probability that a call in the outer zone will move to the inner zone. The complementary probability (the probability of the complementary event that the call will cross the cell boundary) is $1 - q$.

## 7.3 Admission control schemes

In the open literature there are two general approaches to admission control in mobile WiMAX systems as explained in Section 7.1 and summarised here. One of the approaches assumes that the resource environment is fixed and protects only inter-cell handoff calls from forced call interruption due to insufficient resources. The other approach modells the dynamic resource environment and in particular, the non-deterministic resource demand, and gives priority to all active (inter-cell and intra-cell handoff) calls. To evaluate each of the approaches we incorporate them in the system model through two admission control schemes based on the premise stated earlier.

Furthermore, recall that we consider the UGS WiMAX class because we expect to see the AMC effect clearly manifested on system performance for this class. Due to the fact that the UGS class was designed specifically for fixed-size and periodic arrival of packets, which best represents voice traffic (symmetric by nature and sensitive to delay and loss) and that the cut-off concept has been primarily implemented for voice handoff call prioritisation motivated us to use it in our study. In particular, the two schemes designed to represent the concepts of the aforementioned admission control approaches are based on this reservation concept.

We elaborate below on the mathematical description of the non-prioritised scheme as well. The goal of studying the NPS is twofold: (1) we assess the effect of AMC on system performance when no admission control is integrated into the system and (2) we consider the NPS as a reference case when comparing and contrasting the two general approaches described above.

Table 7.1: Acceptance probabilities for new and active calls for the non-prioritised (NPS), traditional (T), and modified (M) schemes

| Call request | **Non-prioritised (NPS)** | **Traditional (T)** | **Modified (M)** |
|---|---|---|---|
| New in zone 1 | $\begin{cases} 1 : C - s_1 \geq 0 \\ 0 : otherwise \end{cases}$ | $\begin{cases} 1 : C - C_g - s_1 \geq 0 \\ 0 : otherwise \end{cases}$ | $\begin{cases} 1 : C - C_g - s_1 \geq 0 \\ 0 : otherwise \end{cases}$ |
| New in zone 2 | $\begin{cases} 1 : C - s_2 \geq 0 \\ 0 : otherwise \end{cases}$ | $\begin{cases} 1 : C - C_g - s_2 \geq 0 \\ 0 : otherwise \end{cases}$ | $\begin{cases} 1 : C - C_g - s_2 \geq 0 \\ 0 : otherwise \end{cases}$ |
| Intra-cell handoff | $\begin{cases} 1 : C - (s_2 - s_1) \geq 0 \\ 0 : otherwise \end{cases}$ | $\begin{cases} 1 : C - C_g - (s_2 - s_1) \geq 0 \\ 0 : otherwise \end{cases}$ | $\begin{cases} 1 : C - (s_2 - s_1) \geq 0 \\ 0 : otherwise \end{cases}$ |
| Inter-cell handoff | $\begin{cases} 1 : C - s_2 \geq 0 \\ 0 : otherwise \end{cases}$ | $\begin{cases} 1 : C - s_2 \geq 0 \\ 0 : otherwise \end{cases}$ | $\begin{cases} 1 : C - s_2 \geq 0 \\ 0 : otherwise \end{cases}$ |

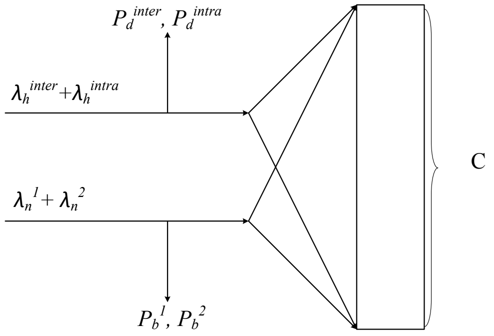

$P_d^{inter}, P_d^{intra}$
$\lambda_h^{inter}+\lambda_h^{intra}$
$\lambda_n^1+\lambda_n^2$
$P_b^1, P_b^2$
C

Figure 7.4: Non-prioritised (complete sharing) scheme: concept

### 7.3.1 Non-prioritised scheme

The non-prioritised (*complete sharing*, *full accessibility*) scheme admits a call request independent of its call class (new or on-going) provided there are enough available slots (see Fig. 7.4 and Table 7.1).

The system state in equilibrium when no admission control is implemented (that is, the NPS is used) can be described by (7.4), where $u_i(x)$ are defined as follows:

$$u_1(x) = u(C - (n_1 + 1) \cdot s_1 - n_2 \cdot s_2) \tag{7.5}$$

$$u_2(x) = u(C - (n_1 - 1) \cdot s_1 - (n_2 + 1) \cdot s_2) \tag{7.6}$$

$$u_3(x) = u(C - n_1 \cdot s_1 - (n_2 + 1) \cdot s_2) \tag{7.7}$$

$$u_4(x) = u_3(x) \tag{7.8}$$

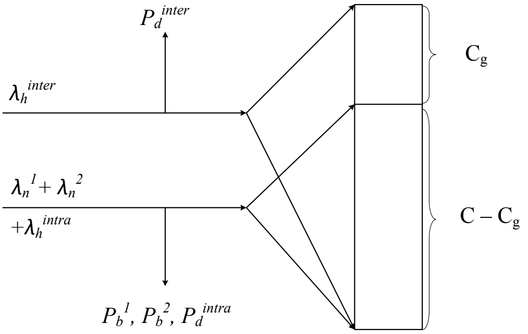


Figure 7.5: Traditional scheme: concept

### 7.3.2 Traditional

The admission control approach that assumes deterministic resource environment is modelled in the following manner. There are no resources explicitly assigned to ongoing calls that experience change in the resource demand but only resources for inter-cell handoffs are reserved. This strategy was incorporated through the cut-off concept (see Fig. 7.5 and Table 7.1): inter-cell handoffs have access to guard slots $C_g$, while a common pool of slots $(C - C_g)$ is accessible to all types of arriving calls (new calls originated in either zone, intra-cell and inter-cell handoffs). We call this scheme traditional cut-off scheme or *traditional* for short.

The system state in this case is given by (7.4), where $u_4(x)$ is defined according to (7.8), and the rest of $u_i(x)$ are given by:

$$u_1(x) = u(C - C_g - (n_1 + 1) \cdot s_1 - n_2 \cdot s_2) \tag{7.9}$$

$$u_2(x) = u(C - C_g - (n_1 - 1) \cdot s_1 - (n_2 + 1) \cdot s_2) \tag{7.10}$$

$$u_3(x) = u(C - C_g - n_1 \cdot s_1 - (n_2 + 1) \cdot s_2) \tag{7.11}$$

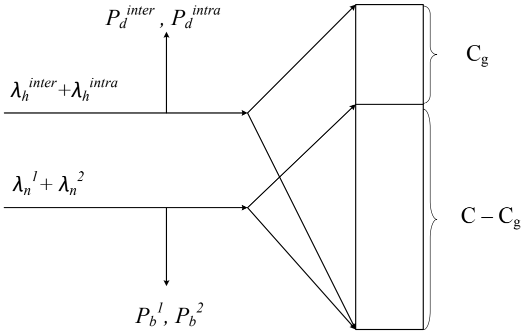

$P_d^{inter}$, $P_d^{intra}$
$\lambda_h^{inter}+\lambda_h^{intra}$
$\lambda_n^1+\lambda_n^2$
$P_b^1$, $P_b^2$
$C_g$
$C-C_g$

Figure 7.6: Modified scheme: concept

### 7.3.3 Modified

The other group of proposals in the open literature considers the dynamic resource demand induced by the radio link variability, and consequently protects all ongoing calls independent of whether these are inter-cell or intra-cell handoffs (see Section 7.1.1). This approach is motivated by the fact that the forced call termination probability of ongoing calls is very obstructive for the mobile user independent of whether the call drop is a result of an inter-cell or intra-cell handoff as the user perceives the outcome—the call interruption—not the factors that lead to such event. We model this concept by introducing a guard capacity $C_g$ reserved for all ongoing calls independent whether arriving from neighbouring cells or requesting additional resources because of presently unfavourable radio link conditions experienced in the serving cell. This general idea in fact corresponds to a strategy proposed in [62, 63].

We call this strategy modified cut-off scheme or succinctly *modified.* Inter-cell handoffs are admitted and call transition to more robust modulation and coding scheme (from inner *1* to outer *2* ring) is allowed as long as there are enough free slots. A new call request is accepted if the number of available slots in the common pool[9] $(C - C_g)$ is not less than the number of slots $s_i$ ($s_1$ in inner zone and $s_2$ in outer zone) that the call requires (see Fig. 7.6 and Table 7.1). In summary, new calls are blocked if there are insufficient slots in the common pool, whereas (inter-cell and intra-cell) handoff calls are dropped if there are not enough slots in the system. Thus, all ongoing calls are given higher priority than newly originating calls. The system state is given by (7.4), while $u_1$, $u_3$, $u_4$ are given by (7.9), (7.11), (7.8) respectively, and $u_2$ is given by:

$$u_2(x) = u(C - (n_1 - 1) \cdot s_1 - (n_2 + 1) \cdot s_2) \tag{7.12}$$

It is worth stressing that the modified scheme accounts for handoffs induced by MCS changes as contrasted by the traditional scheme.

## 7.4 Performance measures

Cellular performance is typically evaluated through the new call blocking probability, forced call termination probability and carried traffic or system capacity. We introduce below the fairness metric. A fairness indicator is generally used in system environments where calls (services) that belong to different classes are susceptible to dissimilar blocking (dropping) probabilities. The fairness concept as discussed earlier has not been investigated in mobile WiMAX systems before.

### 7.4.1 Blocking probability

Since we make a common assumption for Poisson new call arrival traffic, the new call blocking probability in each zone can be determined according to the PASTA property. In particular, the blocking probability in zone $i$ is given by:

$$Pb_i = \sum_{x} P(\vec{n}),$$

[9] Note that the guard resources and the common pool denote amount of system resources.

where

$$x = \left\{ (\vec{n}) \middle| \sum_{j=1}^{j=N} s_j n_j \geq Thr \right\}$$

For the two-zone model, we can specify $Pb_i$ by:

$$Pb_i = \sum_x P(n_1, n_2)$$

where

$$x = \{(n_1, n_2)\, | s_1 n_1 + s_2 n_2 \geq Thr\}$$

and $i = 1, 2$. The threshold $Thr$, depends on the admission control scheme and can be easily deduced from Table 7.1.

The total blocking probability is given by:

$$Pb = p \cdot Pb_1 + (1 - p) \cdot Pb_2,$$

where $p$ is the fraction of the cell area corresponding to the area of the inner zone as defined earlier.

### 7.4.2 Dropping probability

#### Intra-cell dropping probability

When the radio propagation conditions for a given mobile terminal improve (SNR rises above a predefined level), the base station indicates to the mobile station to change the modulation and coding order to a higher one because the same bit error rate can be achieved with higher MCS. When the mobile terminal switches to a higher order MCS due to improvement of the radio link (better SNR), some resources are released. Recall that we assume a simple propagation model, according to which the received power depends chiefly on the distance from the base station. Hence, the mobile terminals distributed in the far end of the cell have the worst radio link in terms of SNR and each call from zone $N$ (zone *2* in the case considered) is assigned the highest amount of resources compared to calls from other zones. Recall also that we model the change in the signal quality of the radio link and the subsequent change in MCS with intra-cell handoffs from one cell zone to a neighbour cell zone. As a result, a handoff from a more outer to a more inner zone cannot lead to a call drop. In effect, call drop can occur when changing to a lower modulation order and concurrently insufficient resources. It means, for the model we adopt, that only a direction of movement from inside towards the edge of the cell is susceptible to call interruption. In the case of two zones, when a call, which is in the immediate vicinity of the base station, moves towards its edge the call can be interrupted because of a dearth of resources.

The dropping probability from the inner to the outer zone is given by the ratio of the dropped intra-cell handoff attempts rate and total intra-cell handoff attempts rate from inner to outer zone:

$$Pd_{intra} = \frac{\sum_y n_1 \cdot P(n_1, n_2)}{\sum_z n_1 \cdot P(n_1, n_2)},$$

where $y$ determines the conditions when intra-cell handoff requests are dropped, and $z$ denotes the system state space. In particular $y$ is given by:

$$y = \{(n_1, n_2) \,| s_1 n_1 + s_2 n_2 \geq= Thr\}$$

The threshold $Thr$ is detemined based on the admission control scheme. The acceptance (rejection) probabilities of the examined schemes are summarised in Table 7.1.

Note, that some authors assume that the intra-cell handoff arrival rate can be approximated by a Poisson arrival process in which case the dropping probability is given by:

$$Pd_{intra} = \sum_y P(n_1, n_2),$$

where $y$ is defined as above. We denote the intra-cell handoff dropping probability by $Pd_1$, i.e., $Pd_{intra} = Pd_1$.

### Inter-cell dropping probability

The inter-cell dropping probability, taking into account the assumption for Poisson inter-cell handoff arrival rate, is given by:

$$Pd_{inter} = \sum_{x} P(n_1, n_2),$$

where $x = \{(n_1, n_2) \,|s_1 n_1 + s_2 n_2 \geq C\}$ for all schemes. We denote $Pd_{inter} = Pd_2$.

## 7.4.3 Forced call termination probability of active calls

The overviewed works on mobile WiMAX admission control do not derive an explicit formulae for the forced call termination probability athough that probability is commonly accepted as a cellular performance measure (the dropping probability indicates the rate with which handoff calls are dropped at a base station not the rate with which calls are interrupted). In [27] a mathematical expression is obtained for the average forced call termination probability when the average number of (intra-cell and inter-cell) handoffs per call is not known. Note though, that the average number of inter-cell handoffs is readily available in mobile cellular networks (because of billing among other purposes) and intra-cell handoffs are known to the base station, as the changes in the modulation and coding scheme are commanded by the base station [2]. Therefore, applying elementary probability theory the forced call termination probability can be found through its complementary probability that the call will not be dropped in either of the visited cells. The forced call termination probability then is given by:

$$Pft = 1 - (1 - Pd_1)^{\alpha_1} \cdot (1 - Pd_2)^{\alpha_2},$$

where $\alpha_1$ and $\alpha_2$ are the average number of intra-cell and inter-cell handoffs per call respectively, and $\alpha = \alpha_1 + \alpha_2$ is the total number of handoffs per call on average.

### 7.4.4 System capacity

An important measure of the efficiency of the handoff priority schemes is the achieved system capacity. The (maximum) capacity is generally defined (see [66] for example) as the maximum offered traffic $A$ or maximum arrival rate $\lambda$ that can be served by the system (i.e., a single cell) such that the QoS requirements—new call blocking probability $Pb$ and forced call termination probability $Pft$—are maintained below their respective levels, that is:

$$A = max\{A : Pb \leq Pb^{max} \cap Pft \leq Pft^{max}\}$$

or equivalently,

$$\lambda = max\{\lambda : Pb \leq Pb^{max} \cap Pft \leq Pft^{max}\}$$

The blocking probability $Pb$ and forced-termination probability $Pft$ are interrelated and increase in one of them in general leads to decrease in the other. The capacity of the system thereby, is limited by the higher probability.

The Telecommunication Standardization Sector of the International Telecommunication Union (*ITU-T*) (formerly the International Consultative Committee for Telephone and Telegraph, or *CCITT*) Series E recommendations cover teletraffic engineering issues in landline and land mobile telephone networks. ITU-T Recommendation E.771 [1] in particular is to the best of the knowledge the only ITU-T recommendation that provides target upper values for the new call blocking $Pb$ and handoff dropping $Pd$[10] probabilities. The ITU-T Recommendation E.771 was approved in 1996 and is still in force. The target values according to Table 4/E.771 and Table 5/E.771 of [1] are $1 \cdot 10^{-2}$ for $Pb$ and $5 \cdot 10^{-3}$ for $Pd$, which values we considered in the numerical evaluations (see Section 7.5). In the light of these considerations, we define system capacity as the maximum new call

[10] The handoff dropping probability (unsuccessful land cellular handover) is defined in [1] as "the probability that a handover attempt fails because of lack of radio resources in the target cell, or because of a lack of free resources for establishing the new network connection".

arrival rate that can be served by the base station (i.e., the system) so that the new call blocking probabilities $Pb_i$ and handoff droping probabilities $Pd_i$ in zone $i$ do not exceed the recommended levels:

$$\lambda_n = max\{\lambda_n : Pb_i \leq Pb^{max} \cap Pd_i \leq Pd^{max}\},$$

that is, the experienced blocking (dropping) probabilities must be lower than the agreed upper values independently of the user location (i.e., the cell zone).

### 7.4.5 Fairness

We emphasised earlier on the importance of the fairness concept to packet-level but also system-level quality of service. We remind the reader that the very notion of mobility suggests that the mobile networks can facilitate support of location-independent service, especially when mobile users of the same service class are charged similarly. This is especially true for the future and IMT-Advanced networks that target both mobility support and ubiquitous coverage—that is, guaranteed service anywhere and anytime regardless of the random wireless channel behaviour. We briefly discuss the understanding of the concept of system-level fairness in publications that contemplate and study this metric and explain how we measure fairness in our work.

Ivanovich *et. al* [53] point out that the goal of admission control schemes should be to attempt to allocate resources in such a way that the blocking and dropping probabilities of calls (half- and full-rate calls in the particular case addressed in [53]) "are equalized as far as possible and at the same time kept to a minimum" [53]. To measure the capability of the network to support uniform system-level performance throughout the service area, a fairness metric $f$ is defined [53] and given by:

$$f = \lg \frac{Pb_i}{Pb_j},$$

where $Pb_i$ and $Pb_j$ are blocking probabilities of calls that belong to the same service class but demand different amounts of resources due to the

Table 7.2: G.723.1 voice coder [92]

| Vocoder | G.723.1 |
|---|---|
| Source bit rate (kbps) | 5.3 |
| Frame duration (ms) | 30 |
| Payload (bytes) (A/I) | (20,0) |

different wireless channels conditions experienced in the cell. The goal is to minimise $f$, while keeping the probabilities low. In effect, complete fairness is achieved when the fairness coefficient $f$ equals 0. Similarly, Epstein and Schwartz [34] stress on the importance of designing fair admission control schemes. The authors set as an objective equalising the blocking probabilities (at most 1% of difference) and maintaining them below certain levels. In our study we argue that as soon as the individual blocking or dropping probabilities of every call class (new or handoff) are below the recommended upper levels, the calls are treated fairly. We consider the fairness coefficient proposed by Ivanovich *et. al* [53] as a measure for absolute fairness but because $f = 0$ in general can be obtained at the price of diminished throughput we establish in our work the condition $\{Pb_i \leq Pb^{max} \cap Pd_i \leq Pd^{max}, i = 1, \ldots, N\}$, where $i$ refers to zone $i$, as sufficient condition for the system to provide fair blocking (dropping) environment to all calls.

## 7.5 Numerical evaluation

### 7.5.1 Scenarios of the experiments

When setting the evaluation scenarios we were guided by [92, 109]. In particular, we consider a mobile WiMAX network with a configuration recommended by the WiMAX Forum [92, 109]: 10 MHz channel with OFDMA symbol time of 102.8 s. and 5 ms. frame, PUSC subchannelization mode and a $DL : UL$ ratio of $2 : 1$ (see Table 4 in [92] for full configuration parameters). The total system capacity is computed according to the guidelines

Table 7.3: Modulation and coding schemes (MCS) used in the numerical evaluations

| Modulation scheme | Spectral efficiency bits/s/Hz | Coding rate | DL (bytes/slot) | UL (bytes/slot) |
|---|---|---|---|---|
| QPSK | 1.5 | 1/2 | 6 | 6 |
| 16-QAM | 3.0 | 1/2 | 12 | 12 |

provided in [92] to obtain numerical results accurate for the WiMAX environment. We assumed that the voice codec specified in Table 7.2 is used[11]. To compute the number of UL (DL) slots per voice user per frame we take into account MAC overheads and packing subheaders [92] but not MAP overhead. The average number of slots assuming enhanced scheduler and parameter values in [92] is given by:

$$\#UL(DL) = \left\lceil \frac{42+6+4}{bytes/slot} \right\rceil$$

where bytes-per-slot rate is determined by MCS according to Table 7.3. The average VoIP packet size with higher layers and enhanced scheduler is taken to be 42 bytes per frame according to [92] and 4 bytes are used for packing subheader [92].

We examine the effect of the adaptive modulation and coding technique on cellular performance under different traffic, signal, and mobility conditions as well as under different admission control disciplines. Different traffic conditions are achieved by varying the offered load and call duration. In order to model the time-varying wireless conditions we vary the rate with which the modulation and coding schemes are changed, i.e., the intra-cell and inter-cell handoff (mobility) rates. The mobility rate changes are achieved by varying the zone dwell times $(1/\eta_i)$. Two different mobility models

[11] In digital communication systems a vocoder is used to encode the compressed digitised speech prior to transmission. In mobile WiMAX no particular vocoder is specified as a recommended one.

and different mobility rates are used to investigate the effect of AMC on system performance. We also investigate the performance of the NPS and the two general approaches reported in the open literature (see Section 7.3) for the aforementioned traffic, signal and mobility conditions.

## 7.5.2 Results and analysis

### Mobility and signal conditions

First, in order to evaluate the effect of AMC on cellular performance under different mobility conditions, we consider two mobility patterns and assess how the AMC impact is manifested under these conditions. In doing so we apply some of the general results of the widely spread zone-based cell model. Later, we evaluate how the switching rate between different MCSs affects system-level performance metrics.

### Mobility model

We consider that mobile terminals are independent and uniformly distributed in the system and move away from the initial position point with equal probability within the $[0, 2\pi)$ range. We differentiate between two mobility models:

- random walk with change of direction (***cd***)

  The model was studied by Zonoozi and Dassanayake [118] and later considered by Cruz-Pérez *et. al* [29] to determine the probability $q$ that a mobile terminal that is located in the outer zone will move into the inner zone of a cell. The assumptions with regard to the model considered in [29] (in addition to those listed above) are: (i) mobile terminals move in straight line with constant speed along a given distance interval (after the selected final point is reached a new direction and speed from a uniform distribution are chosen), and (ii) the probability of the variation of the direction follows a uniform distribution limited in the range of ±180 relative to the current direction of the mobile. Computer simulations are developed in [29] to calculate the probability $q$ using the random walk mobility model. Based on paramterisation of the obtained simulation results the parameter $q$ is defined in [29] by:

$$q = -0.0427675982p + 0.539651632p^{0.218298231}, \qquad (7.13)$$
$$0 \leq p \leq 1,$$

where $p$ is the proportion of the inner zone to the cell area as defined earlier.

- random walk with constant direction (***rw***)

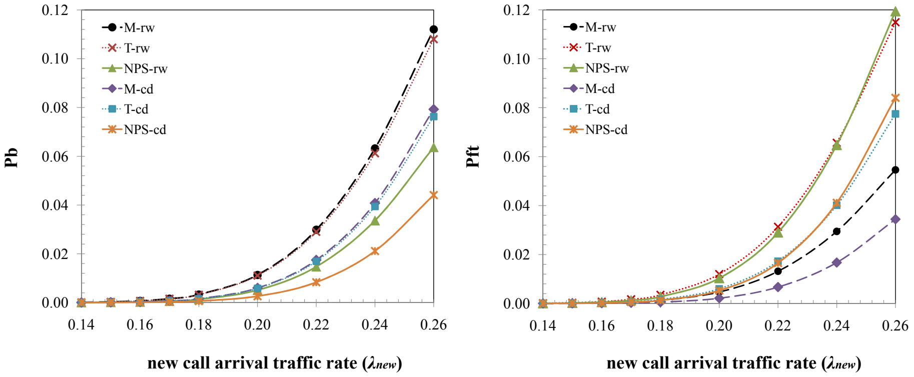
  Like in random walk **cd**, the mobile terminals move in straight line but there is no change in the direction of movement. The probability of moving to the inner zone starting from the outer zone is determined geometrically by the ratio of the angle that encompases all directions that will lead the mobile into the inner zone to all possible directions (i.e., $2\pi$)[12] [29, 97]. The probability $q$ in this case is given by:

$$q = \frac{1}{\pi(1-p)} \left( -\frac{p\pi}{2} + \sqrt{p(1-p)} + \arcsin(\sqrt{p}) \right), \qquad (7.14)$$
$$0 \leq p < 1$$

The random walk mobility model is often used in mobility studies, which motivated us to consider it in our work.

The numerical results ploted in Fig.7.7 clearly show that the effect of AMC on system performance depends on the mobile environment. When mobiles follow a random walk **cd** mobility model the blocking and dropping probabilities observed are lower compared to those obtained when there is no change in the direction of movement. This conclusion can be easily understood taking into account the following. The probability that a user in the outer zone will move into the inner zone is higher for the random walk mobility model (see Fig.7.8). Consequently, the rate with which

[12] The derivation of (7.14) is straightforward, see for instance [29, 97].



**(a)** New call blocking probability

**(b)** Forced call blocking probability

Figure 7.7: System-level performance for different offered traffic load. Two mobility cases are considered: (i) random walk with change of direction (**cd**) and (ii) random walk with constant direction (**rw**) movement.

mobiles move to inner zone will be higher than that of the random walk with constant direction (denoted ***rw*** in the figures) mobility model. An intra-cell handoff from outer to inner zone frees resources (in our scenario: $s_2 - s_1 = 4$ slots are released) in contrast to the intra-cell handoff from outer to inner zone, which requires the same number of additional resources (4 slots). There will be more free resources in a system with random walk (**cd**) compared to the random walk with constant direction (**rw**) mobility model and as a result the blocking and dropping probabilities will be lower for the former.

Note that in the rest of the numerical examples the random walk with constant direction mobility model is used.

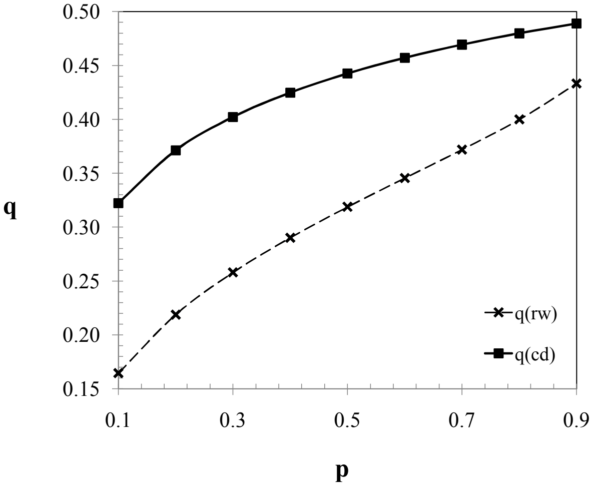
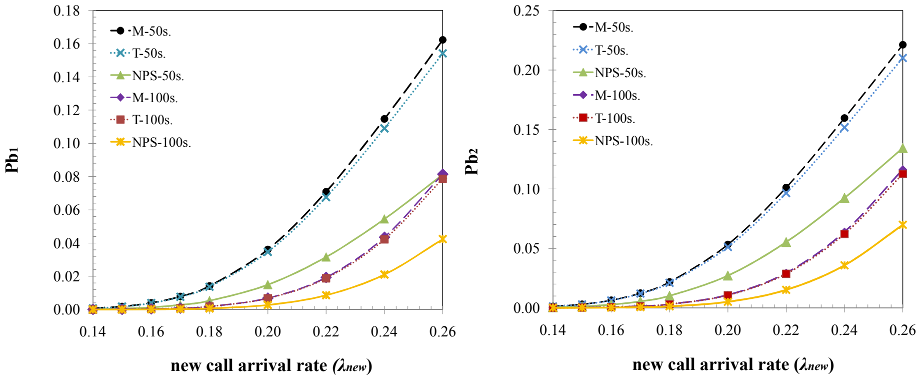


Figure 7.8: Probability **q** that a mobile terminal located in the outer zone will move into the inner zone versus the proportion **p** of the inner zone. Two mobility models are considered: (i) random walk with change of direction (**cd**) and (ii) random walk with constant direction (**rw**) mobility patterns.

### Mobility rate (dwell time)

The mobility rate—the rate, with which the mobile terminals cross zone boundaries—is used to model the time-varying radio channel conditions and as a result the rate with which the MCS of ongoing calls is changed. The mobility rate $\alpha$ on the other hand is controlled through the cell dwell time $\eta$ and call holding time $\mu$ (see for instance [18,58]). Recall from Section 7.2.3 that $\alpha = \eta/\mu$. A larger $\eta/\mu$ modells higher user mobility ($\alpha$ larger than 1), and vice versa, samller $\eta/\mu$ reflects lower user mobility. The intra-cell mobility is modelled following a similar reasoning for the relation between the zone dwell time and call duration time.

We assume that each zone occupies half of the cell area, i.e., $p = 0.5$. The

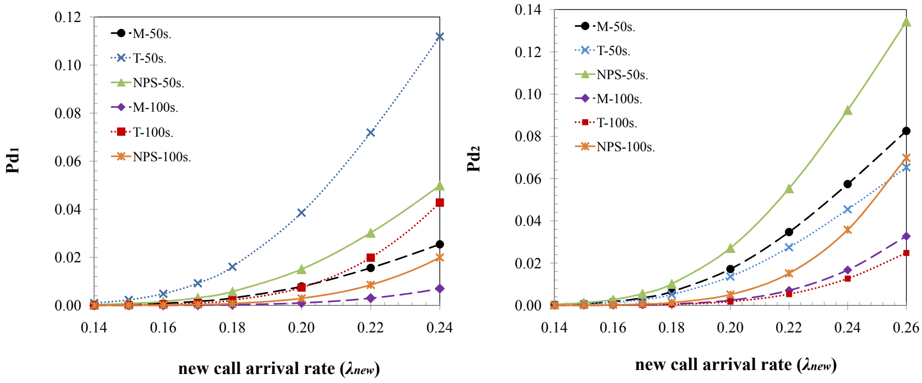


**(a)** zone 1

**(b)** zone 2

Figure 7.9: New call blocking probabilities for different mobility rate and offered traffic load conditions. Random walk with constant direction mobility model is assumed. Two mobility rate cases are considered: (i) zone dwell time $1/\eta_i = 50$ s. and (ii) zone dwell time $1/\eta_i = 100$ s.

mobility model is random walk with constant direction. The call duration is $1/\mu = 180$ s. Two cases are examined: (i) mean zone dwell time $(1/\eta_i)$ equal to 50 s. and (ii) mean zone dwell time $(1/\eta_i)$ equal to 100 s. The results plotted in Fig. 7.9 and Fig. 7.10 show that the lower the residence time in the signal-defined cell zones, the higher the blocking (dropping) probabilities. That is, if the frequency with which a call experiences changes in the MCS is increased, then the probability of a call drop is also increased. It means that, for the assumed analytical model, the two intra-cell handoff events (moving from inner to outer and moving from outer to inner zone) do not cancel each other but more frequent MCS changes lead to higher dropping and consequently forced call termination probabilities. The rate of MCS adaptation has a similar effect on the blocking probabilities namely, for higher MCS change rate higher blocking is observed.



**(a)** zone 1

**(b)** zone 2

Figure 7.10: Dropping probabilities for different mobility rate and offered traffic load conditions. Random walk with constant direction mobility model is assumed. Two mobility rate cases are considered: (i) zone dwell time $1/\eta_i = 50$ s. and (ii) zone dwell time $1/\eta_i = 100$ s.

### Traffic conditions

### Offered load

The results plotted in Fig. 7.7, Fig. 7.9, and Fig. 7.10 demonstrate that the heavier the traffic load the more pronounced is the effect of AMC on system-level peroformance. The higher the arrival traffic rate under otherwise equal conditions (system capacity, call duration, residence time, etc.) the higher the probability that the resources will be busy and consequently the higher the probability that the call requests will be blocked or dropped. Therefore, except for light offered traffic, the random wireless channel behaviour and as a result the time varying resource demand shall be addressed by resource reservation or call prioritisation techniques that can guarantee the continuity and quality of the ongoing calls.

### Call duration

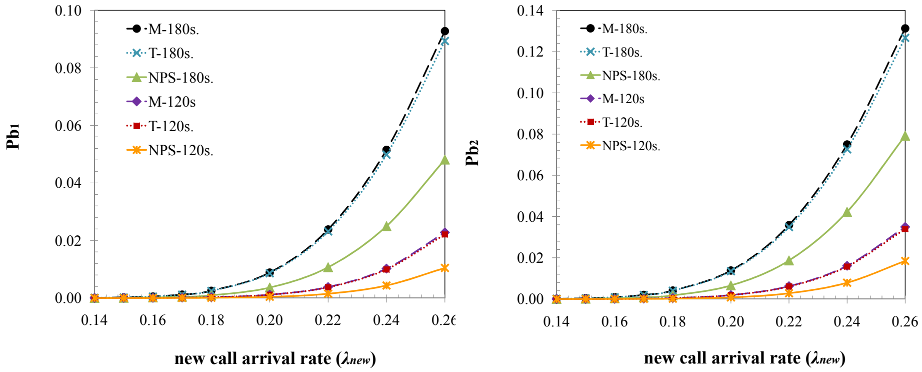
The call duration increases the blocking and dropping probabilities as illustrated by Fig. 7.11 and Fig. 7.12. The blocking (dropping) probabilities are higher because the resources are seized for longer time and therefore the probability that there will be free resources when a new (handoff) call request arrives is lower. Consider for instance the inter-cell dropping probability. When the call holding time is longer but the mean cell dwell time remains constant, the probability that the call will be accomplished before requesting an inter-cell handoff will decrease; that is, the inter-cell handoff rate for otherwise equal zone dwell time and system resources will be increased. The latter means heavier ongoing traffic load, which as demonstrated above worsens system-level performance.

The new blocking probability in the inner zone is less severely affected by the longer call duration than the call blocking probability in the outer zone as the resources needed in the good signal quality zone are fewer compared to the other zone. Likewise, the intra-cell dropping probability is lower than the inter-cell dropping probability because of the fewer slots required to successfully handing off an intra-cell call request compared to an inter-cell handoff request.

### Admission control approaches

In general, for high traffic, mobility, and frequently changing signal conditions the non-prioritised scheme leads to concurrent increase in the new call blocking and active call dropping probabilities as discussed before.

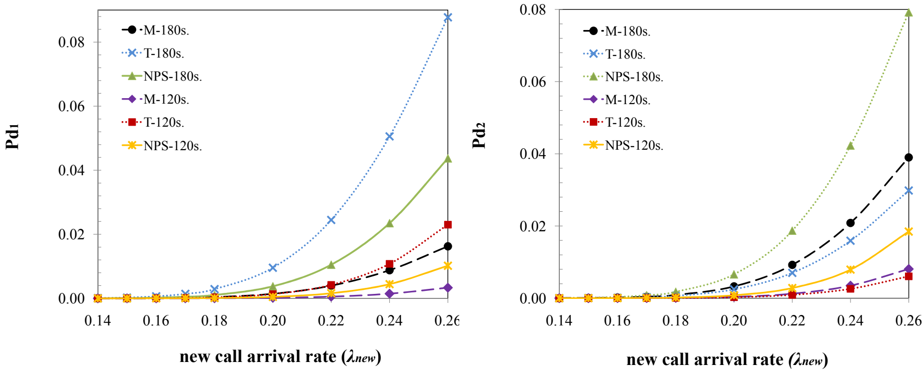
When compared to the NPS the traditional and modified schemes (denoted T and M schemes in the figures) lead to higher blocking probabilities in both zones, which was expected because both implement the guard channel concept. The new call arrivals have access to a portion of the resources (i.e., to the common pool) rather than the entire system capacity as it is in the case of the non-prioritised scheme, which yields more intensive new call blocking rate. The total blocking probabilities for the traditional and modified schemes are practically the same but larger than $Pb$ for the NPS.



**(a)** zone 1 **(b)** zone 2

Figure 7.11: New call blocking probabilities for zone dwell time with mean $\eta^{-1} = 90$ s., different call duration and offered traffic load conditions. Two call holding times $(1/\mu)$ are considered: (i) $1/\mu = 120s$ and (ii) $1/\mu = 180s$.

The intra-cell handoff dropping probability of the traditional guard channel scheme is significantly higher than that of the reference case (that is, $Pd_1$ of the NPS). This is due to the restricted access to the system resources; namely, intra-cell handoffs are admitted only if there are enough free slots in the common pool of resources. The modified guard channel scheme achieves better intra-cell dropping probability $Pd_1$ than both the traditional scheme and the NPS as the handoff calls are always prioritised regardless of their class (i.e., intra-cell or inter-cell). The inter-cell dropping probability $Pd_2$ for the traditional scheme is lower than that for the NPS as expected. The modified scheme leads to slightly higher inter-cell dropping probability than the traditional scheme because the guard slots are reserved for intra-cell and inter-cell handoffs rather than exclusively for inter-cell handoff requests but is much lower than the $Pd_1$ of the NPS. It is noteworthy that the intra-cell dropping probability for the traditional scheme is of orders of magnitude



**(a)** zone 1

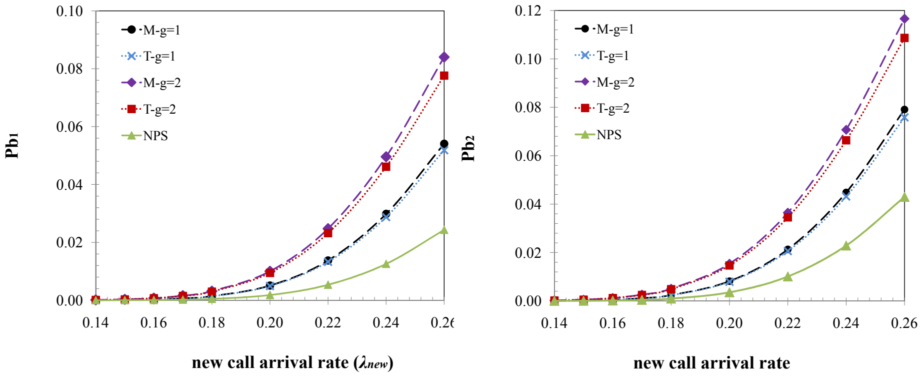
**(b)** zone 2

Figure 7.12: Dropping probabilities for zone dwell time with mean $\eta^{-1}$ = 90 s., different call duration and offered traffic load conditions. Two call holding times $(1/\mu)$ are considered: (i) $1/\mu = 120s$ and (ii) $1/\mu = 180s$.

higher than that of the NPS and the modified scheme. As a result, for all the studied scenarios the forced call termination probability[13] for the traditional scheme considerably exceeds that of the modified scheme and importantly it is higher than the non-prioritised scheme.

These observations lead to the conclusion that the dynamic resource demand in mobile WiMAX with AMC must be addressed by admission control strategies, otherwise unacceptable cellular performance can be observed. In particular, admission control schemes that do not model the adaptive resource demand induced by the time varying radio channel conditions in a cell can, in fact, worsen system-level performance compared to the reference case when no admission control is used. Note however, that we do not expect an identical effect—poor cellular performance when the traditional

[13] The forced call termination probability was calaculated for scenarios when there is at least one inter-cell handoff and one intra-cell handoff.



**(a)** zone 1

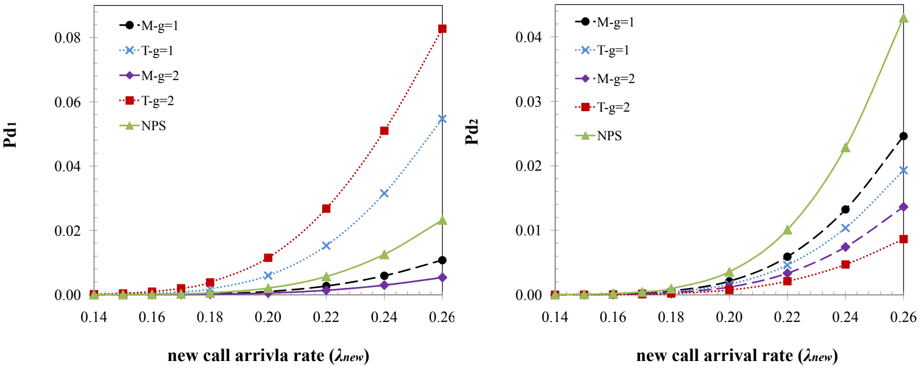
**(b)** zone 2

Figure 7.13: Performance of the Traditional (T) and Modified (M) schemes for different guard capacities: new call blocking probabilities. Two guard capacity cases are considered: (i) guard capacity $g = 1$ and (ii) guard capacity $g = 2$.

approach is implemented compared to a system with complete sharing of resources—to be observed (it will be at least less prominent) with the other WiMAX scheduling classes (i.e., ertPS, rtPS, nrtPS, and BE).

We also examined the effect of AMC on system performance when the guard capacity takes different portions of the whole system capacity. The aforedescribed blocking and dropping trends are further strengthen when the number of reserved slots is doubled as illustrated by Fig. 7.13 and Fig. 7.14.

Individual blocking and dropping probabilities are plotted in the figures (see Fig.7.7 to Fig.7.12) insted of the total blocking and forced call termination probabilities to disclose the effect of the adaptive modulation and coding technique on system-level fairness under different admission control policies. The results illustrate that none of the admission approaches achieves absolute fairness (recall the fairness index $f$ proposed by Ivanovich [53]).



**(a)** zone 1

**(b)** zone 2

Figure 7.14: Performance of the Traditional (T) and Modified (M) schemes for different guard capacity conditions: dropping probabilities. Two guard capacity cases are considered: (i) guard capacity $g = 1$, and (ii) guard capacity $g = 2$.

Consider for instance the modified scheme. The blocking probabilitites in the inner and outer zone are of a different order because although the new calls can access the same amount of resources, the resources they demand depend on the zone where the calls are initiated. Likewise, the dropping probabilities depend on the zone because all active calls have access to all system resources but require different amount of resources to be served. System performance is discussed below for conditions that satisfy the fairness requirement as we specified it earlier; that is, when the blocking (dropping) probabilities in each zone are below the prescribed upper levels.

As noted above, the numerical results show that the modified scheme achieves much better handoff performance than the NPS but at the prise of much higher blocking probability. We focus below on the system capacity attained by these two schemes. Recall that due to required blocking and dropping fairness, the maximum arrival rate is defined as the arrival rate that can be served by the system while keeping the individual blocking and dropping probabilities below predefined levels. Our numerical results show that the system capacity for the modified and non-prioritised schemes are practically the same ($\lambda_n$ about 0.2 when $\{Pb_1 \cap Pb_2 \leq 1 \cdot 10^{-2}\}$ $\cap\{\mathrm{Pd}_1 \cap Pd_2 \leq 5 \cdot 10^{-3}\}$). The main difference between the performance of the two schemes is that with the NPS, system capacity is limited by the dropping probability as contrasted by system capacity that is limited by the blocking probability for the modified scheme. Therefore, it can be concluded that there is no gain in capacity but better handoff performance can be achieved with the modified scheme. Specifically, the modified scheme keeps $Pd_i$ much below the recommended levels, while $Pb_i$ is just below or equal to the recommended upper $Pb^{max}$ value. The complete sharing scheme maintains $Pd_i$ about the predifined upper level, while $Pb_i$ is much below $Pb^{max}$. It is noteworthy that the guard channel scheme has been of much practical interest to the conventional mobile cellular systems not only because of its simplicity but also because the introduction of a guard capacity resulted in a slight increase in the blocking probabilty but significant decrease in the dropping probability (the relative increase in $Pb$ is in general much smaller than the relative decrease in $Pd$). It is important to contrast this general result with the observed system performance plotted in Figs. 7.7 to Fig. 7.12 namely, the increase in the blocking probability is of the same order as the decrease in the dropping probability. The reason behind these results is that the guard capacity affects more seriously the slot-hungry (underpriviledged) calls, which is reflected in the values of $Pb_2$ and eventually in the total new call blocking probability $Pb$.

The observed trends lead to conclusions that could serve as guidelines in mobile WiMAX admission control design namely,

- When the adaptive modulation and coding technique is implemented in the system, the fluctuating radio channel and the resulting dynamic resource demand do not have important effect on the blocking and dropping probabilites only for light offered traffic conditions (in the numerical studies arrival rate of $\lambda_n$ below 0.16 call per time unit can be considered as light load). For the rest of traffic conditions the time-varying radio channel conditions might lead to system performance deterioration expressed in terms of increased $Pb$ and $Pft$. We also examined how system performance is affected by the rate with which the signal quality of the radio link varies. The results showed that for higher radio channel variability under otherwise equal conditions, the experienced blocking (dropping) probabilities are higher. In summary, for moderate to heavy traffic load, quickly varying signal conditions, as well as more deterministic mobility patterns (random mobility versus random mobility with constant direction), the effect of AMC on the blocking and dropping probabilities of a constant bit rate service, is manifested in worse performance.

- When the cellular system fairness in terms of blocking and dropping probabilities is not addressed the performance of the system is determined by the higher blocking (dropping) probability of the most slot-consuming calls. Recall that we have established as a requirement for system-level fairness that all individual blocking (dropping) probabilities must be kept below predefined upper levels. However, when techniques that equalise these probabilities are not incorporated into the system, cellular performance is limited by the performance measures of the calls with the highest slot demand. Therefore, in such case it is sufficient to design the admission control policy so that the new call blocking probability in the outer ring and the inter-cell handoff dropping probability are kept below the upper levels. The rest of blocking (dropping) probabilities are inherently lower than those aforementioned.

- It was shown that simple modification of the classical cut-off concept brings some advantages compared to system performance without admission control. The advantages are in terms of dropping probabilities considerably lower than the upper level, which is of benefit to the mobile user[14].

[14] Forced call termination is much more obstructive than blocking probability, therefore it is preferable for the same system capacity to achieve lower dropping than blocking probability. Recall that according to ITU-T Recommendation E.771 “The probability of an unsuccessful handover is a critical parameter in a cellular system, as an unsuccessful handover affects a call already in progress.”

### 7.5.3 Conclusions

The numerical results, derived from the analytical zone-based cell model of a mobile WiMAX system, show that the adaptive modulation and coding technique affects system performance. In particular, the total new call blocking probability and forced call termination probability of a constant bit rate calls are increased when the radio channel conditions are quickly varying and the offered load is moderate to heavy. Importantly, the results demonstrate that the random resource demand due to the adaptive modulation and coding technique must be addressed by admission control to avoid frequent call interruption, that is, to support the continuity and quality of the calls.

Furthermore, we studied the effect of AMC on the blocking and dropping probabilities under the two basic admission control approaches proposed in the literature.

To model the class of admission control schemes that assume fixed resource allocation and examine the impact of such hypothesis on system performance we devised a scheme that we denoted traditional. It was demonstrated that system performance under such approach is in fact worse than system performance when no admission control is implemented. Generalising the latter outcome, we assert that the specifics of the technology cannot be ignored and should be taken into consideration when developing and implementing admission control algorithms.

To model the class of admission control algorithms that address the dynamic resource environment of mobile WiMAX systems with AMC we incorporated the modified scheme. The scheme achieves lower dropping probabilities and consequently a lower forced call termination probability but at the prise of a higher blocking probability than the non-prioritised scheme. Furthermore, the modified approach does not retain the advantageous property that the cut-off scheme has in conventional networks—guard channel(s) yield small increase in the new call blocking probability but considerable decrease in the dropping probability. This property is not kept in mobile WiMAX mainly due to variablity in the call's resource demand.

The numerical results clearly showed important differences in the blocking and dropping probabilities of calls belonging to the same service (*voice*) and call (either *new* or *handoff*) class but being served in different modulation and coding scheme zones. The fairness problem was not studied in previous works [17,62,63,101–103,107] despite the relevance of the need for providing uniform service to the (mobile) users [34, 52, 53, 60]. We conjecture that fair resource usage (that is, uniform service throughout the cell area) will improve cellular performance in terms of blocking and dropping probabilities and (or) system capacity compared to a network that lacks system-level fairness.

## 7.6 Concluding remarks

The focus in this chapter was on the dynamic resource environment typical for the next generation wireless networks. The overall contribution of this research is that the effect of the adaptive modulation and coding mechanism on system performance (i.e., new call blocking and forced call termination probabilities) was evaluated, which practical question has not been investigated in previous works. This research also identified the system-level unfairness problem, which has not been studied in the relevant WiMAX literature.

We based our study on an analytical model that captures the dynamics of resource usage induced by changes in the signal quality of the wireless channel and user mobility. We assessed the effect of the adaptive modulation and coding technique on the dropping, forced call termination and new call blocking probabilities in mobile WiMAX networks, as well as on system-level performance fairness considering voice service. Two admission control schemes were incorporated into the analytical model to model the two general admission control approaches proposed in the mobile WiMAX literature. The class of admission control approaches that do not model the dynamics of the resource demand (i.e., assume fixed resource allocation as in the traditional networks) was modelled by the traditional cut-off concept, whereas the other class of approaches that address the changes in the radio link was modelled by the modified cut-off concept. Thus, we could examine the pure effect of AMC for system without call prioritisation but also determine the validity of the hypothesis for deterministic resource conditions.

The anayitical results for a homogeneous (in terms of traffic rates, cell capacity and signal zones) system and voice service showed that under moderate to heavy traffic load, quickly varying signal conditions, as well as more deterministic mobility patterns (random walk versus random walk with constant direction), the effect of AMC on blocking and dropping probabilities is manifested in worse system performance. Furthermore, it was demonstrated that the traditional approach can worsen cellular performance in comparison with the case when no admission control policy is implemented into the system. Therefore, the system shall prioritise not only active calls that arrive from neighbouring cells (i.e., inter-cell handoff calls) as it is in the systems with fixed allocation but also active calls that do not change the serving cell but experience increase in the resource demand due to deterioration of the signal quality of the radio link. The other major impact of the non-deterministic resource allocation is on system-level fairness because the more resource-hungry active (new) calls are susceptible to higher forced call termination (new call blocking) probability than the less resource-hungry calls. That is, the mobile WiMAX systems are inherently unfair in terms of blocking and dropping probabilities and explicit measures shall be incorporated in order to provide uniform quality of service within the service area.

# 8 Conclusions

Field trials, started in the late 90s, in functional mobile cellular systems have revealed that several of the teletraffic variables that describe these networks are not exponentially distributed contrary to the common Poisson or memoryless assumptions, which had governed the research in admission control previously. Motivated by these empirical results, many researchers centred on further characterising analytically, empirically, or through simulation the real probabilistic nature of these processes (see Chapter 2). The effect of the traditional exponential hypothesis on system performance has been investigated as well. Posteriori, various research groups carried out studies that aimed at developing an analytical framework for system performance evaluation for non-conventional conditions such as generally distributed call and channel holding time. Although the statistical studies have been initiated under the umbrella of system performance evaluation and have been aimed at better understanding the processes observed in mobile cellular networks and eventually at system optimisation, there have not been contributions that actually used these results for improving system performance by means of admission control design. Therefore and because of the scientific conjecture that such investigations would yield interesting results and solutions, in Chapter 3 we explored the main outcomes of this intensive research field from the perspective of admission control. In particular, we devised a teletraffic based scheme that controls the acceptance of new call arrival traffic by estimating the future system occupancy state. The estimations are based on the statistical profile of the channel holding time random variable. Further, the metrics implemented in the admission control scheme are easily obtained in the system. The main advantages of the devised scheme are (i) implementation simplicity, which is an important practical requirement, (ii) fast execution, which allows for quick admission control response, (iii) a continuous working interval, which provides mobile operators with (more compared with the classical cut-off scheme) freedom to make a trade-off between quality of service and efficient resource use. Both analytical and simulation approaches were used to evaluate the performance of the developed scheme. The mathematical analysis was carried out for traditional exponential assumptions. A simulation pure performance model was developed for scheme's assessment for non-trivial conditions. In Chapter 4 the proposed scheme was evaluated for high handoff rates, overlapping areas, and non-Poisson arrival traffic flows, which conditions matched measured data from real, live mobile cellular networks. In Chapter 5 the hypothesis for channel holding time distribution availability at the base station was relaxed. A feasible approximation was suggested and its validity was studied under general channel holding time conditions. The central result from the research reported in the first part of the thesis is that the statistical profile of the teletraffic variables that describe mobile cellular networks can be advantageously implemented in admission control for improving system performance and user satisfaction.

The wireless technologies underwent remarkable development in the last few decades because the proliferation of the mobile communications accentuated the dearth of the radio resources and there was an imperative need for more efficient radio resource solutions. However, the mobile traffic demand and the communication technologies have been evolving in parallel. Therefore, it is not surprising that the capacity of the mobile cellular networks has continued to be scarce and even nowadays, despite the considerable engineering achievements in the field, it is insufficient. Consequently, radio resource management and admission control in particular keep their essential role in providing quality of service in an environment characterised by radio resource scarcity. The novel techniques that were successfully deployed in the more advanced wireless systems and adopted by the modern broadband wireless technologies (see Chapter 6), such as the spectrum-efficient link adaptation, introduced new random variables that drastically changed the traditional mobile cellular system model. The adaptive adjustment of the transmission rate to the random radio link behaviour for instance, leads to cell capacity and call resource demands that are not deterministic but time-varying. The dynamic resource environment complicates the system quality of service provisions as well as system analysis. Hence, it is of much practical interest to investigate the new context within which admission control must be devised and executed. We focused our research effort on modelling and analysing the effect that the adaptive modulation and coding technique has on system performance and finally on admission control design. In Chapter 7 this question was investigated mathematically. A zone-based cell model was used to model the randomness of the radio channel and the consequent non-deterministic resource demand of a streaming service with constant bit rate and strict delay requirements. It was shown that the dynamic resource conditions can yield system-level unfairness and negatively impact the voice connections that experience unfavourable radio conditions. Moreover, we showed that if the admission control is not adapted to the actual system environment—in particular we studied the traditional cut-off scheme performance in mobile WiMAX networks—it can lead to worse system performance compared to the case when no admission control scheme is used. Importantly, we concluded that the admission control design must address the random resource demand in order to avoid frequent call interruptions, that is, to support continuity and quality of calls.

In the scope of the reported research work there are several lines that we consider worth investigating further. A natural continuation of our investigations is the mathematical analysis of the proposed MRT scheme for non-trivial scenarios. It is our conviction that although not straightforward the analysis is feasible and could bring new insights into the functionality of the algorithm and thus, can suggest new ways of exploring the inherent statistical properties of the teletraffic random variables. Novel and more efficinet algorithms can be devised as a result. Furthermore, the exploitation of statistical estimates in a more diverse teletraffic environment, as the one proper for the broadband mobile cellular networks, is expected to bring additional system performance improvements compared to the conventional monoservice environment. The latter conjecture is motivated by the flexibility of the elastic traffic flows in meeting their quality of service requirements. Another logical continuation of our work is to devise admission control scheme that takes into account the main conclusions drawn from the research reported in Part II of the thesis. As a further step, it can be elaborated on an advanced system model that incorporates more technological features such as HARQ and MIMO system. It will be interesting to examine their effect on system capacity and eventually on admission control. This research can be further extended by a simulation study that on one hand validates the analytical results and on the other hand evaluates system performance for conditions that are difficult to model mathematically.

In more broader scope, there are several research questions in the context of admission control that are worth considering. The relay stations and femtocells, which have been recently incorporated in the advanced mobile wireless standards, are designed to alleviate the radio resource problem but introduce additional variables to the system model. The latter impose new interrogatives to admission control. The adoptation of green communication concepts, the co-existance of various wireless communication systems based on different standards, the diversification of services, that is, the heterogeneous radio, system, and service conditions, are some of the factors that are defining the new environment and therefore challanges to admission control.

# Bibliography

[1] "Network grade of service parameters and target values for circuit-switched land mobile services," *International Telecommunication Union – Telecommunication Standardization Sector (ITU–T) (formerly the International Consultative Committee for Telephone and Telegraph, or CCITT) Recommendation E.771 (previously CCITT Recommendation)*, Oct. 1996.

[2] "Part 16: Air interface for Broadband Wireless access systems, (Revision of IEEE Stdandard 802.16-2004 and consolidates material from IEEE Std. 802.16e-2005, IEEE 802.16-2004/Cor.1-2005, IEEEE Std. 802.16f-2005, and IEEE Std. 802.16g-2007)," *IEEE standard*, Jan. 2009.

[3] "Cisco visual networking index: global mobile data traffic forecast update, 2010-2015," *Cisco White Paper*, pp. 1–27, Feb. 2011.

[4] V. Aalo and G. Efthymoglou, "Performance analysis of generalized handover model for cellular networks," in *IEEE Sarnoff Symposium*, April 2009, pp. 1–5.

[5] M. Ahmed, "Call admission control in wireless networks: a comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 7, no. 1, pp. 49 –68, qtr. 2005.

[6] M. Ahmed, H. Yanikomeroglu, and S. Mahmoud, "Fairness enhancement of link adaptation techniques in wireless access networks," in *58th IEEE Vehicular Technology Conference (VTC-Fall 2003)*, vol. 3, 2003, pp. 1554–15 573.

[7] ——, "Fairness of link adaptation techniques in broadband wireless access networks," in *59th IEEE Vehicular Technology Conference (VTC-Spring 2004)*, vol. 4, May 2004, pp. 1944–1948.

[8] A. Alfa and W. Li, "PCS networks with correlated arrival process and retrial phenomenon," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 630 – 637, Oct. 2002.

[9] A. O. Allen, *Probability, Statistics, and Queueing Theory with Computer Science Applications.* Academic Press, Inc., 1990.

[10] F. Barcelo, "Statistical properties of silence gap in public mobile telephony channels with application to data transmission," in *IEEE International Conference on Communications (ICC 2001)*, vol. 7, 2001, pp. 2011–2015.

[11] ——, "Performance analysis of handoff resource allocation strategies through the state-dependent rejection scheme," *IEEE Transactions on Wireless Communications*, vol. 3, no. 3, pp. 900–909, May 2004.

[12] F. Barcelo and S. Bueno, "Idle and inter-arrival time statistics in public access mobile radio (PAMR) systems," in *IEEE Global Telecommunications Conference (GLOBECOM'97)*, vol. 1, Nov. 1997, pp. 126–130.

[13] F. Barcelo and J. Jordan, "Channel holding time distribution in cellular telephony," *Electronics Letters*, vol. 34, no. 2, pp. 146–147, Jan. 1998.

[14] ——, "Channel holding time distribution in public telephony systems (PAMR and PCS)," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1615–1625, Sep. 2000.

[15] F. Barcelo and J. Sanchez, "Probability distribution of the inter-arrival time to cellular telephony channels," in *IEEE 49th Vehicular Technology Conference (VTC'99)*, vol. 1, jul 1999, pp. 762–766.

[16] V. Bolotin, "Modeling call holding time distributions for CCS network design and performance analysis," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 3, pp. 433–438, Apr. 1994.

[17] T. Chahed and C. Tarhini, “Impact of mobility on the performance of data flows in OFDMA-based IEEE802.16e systems,” in *IEEE International Symposium onWireless Communication Systems (ISWCS ’08)*, 2008, pp. 648–652.

[18] I. Chlamtac, Y. Fang, and H. Zeng, “Call blocking analysis for PCS networks under general cell residence time,” in *IEEE Wireless Communications and Networking Conference (WCNC’99)*, 1999, pp. 550–554.

[19] E. Chlebus, “Empirical validation of call holding time distribution in cellular communications systems,” in *15th Int. Teletraffic Congress (ITC)*, Elsevier, Ed., 1997.

[20] E. Chlebus and W. Ludwin, “Is handoff traffic really Poissonian?” in *4rth IEEE International Conference on Universal Personal Communications*, Nov. 1995, pp. 348–353.

[21] T. Christensen, B. F. Nielsen, and V. B. Iversen, “Phase-type models of channel-holding times in cellular communication systems,” *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 725–733, May 2004.

[22] A. Corral-Ruiz, A.L.E. Rico-Paez, F. Cruz-Perez, and G. Hernandez-Valdez, “On the functional relationship between channel holding time and cell dwell time in mobile cellular networks,” in *IEEE Global Telecommunications Conference (GLOBECOM’10)*, Dec. 2010, pp. 1–6.

[23] A. Corral-Ruiz, F. Cruz-Perez, and G. Hernandez-Valdez, “Channel holding time in mobile cellular networks with generalized Coxian distributed cell dwell time,” in *IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC’10)*, Sept. 2010, pp. 2348–2353.

[24] ——, “Teletraffic model for the performance evaluation of cellular networks with hyper-erlang distributed cell dwell time,” in *IEEE 71st Vehicular Technology Conference (VTC-Spring 2010)*, May 2010, pp. 1–6.

[25] D. Cox and D. Reudink, "Some effects on channel occupancy of limiting the number of available servers in small cell mobile radio systems using dynamic channel assignment," *IEEE Transactions on Communications*, vol. 27, no. 8, pp. 1224–1226, Aug. 1979.

[26] F. Cruz-Perez, G. Hernandez-Valdez, and L. Ortigoza-Guerrero, "Performance evaluation of mobile wireless communication systems with link adaptation," *IEEE Communications Letters*, vol. 7, no. 12, pp. 587–589, Dec. 2003.

[27] ——, "Performance evaluation of mobile wireless communication systems with link adaptation," *IEEE Communications Letters*, vol. 7, no. 12, pp. 587–589, Dec. 2003.

[28] F. Cruz-Perez, A. Seguin-Jimenez, and L. Ortigoza-Guerrero, "Residence time distribution in mobile cellular systems with link adaptation," in *IEEE Wireless Communications and Networking Conference(WCNC'04)*, vol. 4, 2004, pp. 2387–2392.

[29] F. Cruz-Perez, J. Vazquez-Avila, G. Hernandez-Valdez, and L. Ortigoza-Guerrero, "Link quality-aware call admission strategy for mobile cellular networks with link adaptation," *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2413–2425, Sep. 2006.

[30] S. Dharmaraja, K. S. Trivedi, and D. Logothetis, "Performance modeling of wireless networks with generally distributed handoff interarrival times," *Computer Communications*, vol. 26, no. 15, pp. 1747–1755, 2003.

[31] J. Diederich and M. Zitterbart, "Handoff prioritization schemes using early blocking," *IEEE Communications Surveys Tutorials*, vol. 7, no. 2, pp. 26 – 45, quarter 2005.

[32] G. Efthymoglou, S. Pattaramalai, and V. Aalo, "Call completion probability with generalized call holding time and cell dwell time distributions," in *IEEE 69th Vehicular Technology Conference (VTC-Spring 2009)*, Apr. 2009, pp. 1–5.

[33] S. Elayoubi, T. Chahed, and G. Hebuterne, “Mobility-aware admission control schemes in the downlink of third-generation wireless systems,” *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 245–259, Jan. 2007.

[34] B. Epstein and M. Schwartz, “Predictive QoS-based admission control for multiclass traffic in cellular wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 523–534, Mar. 2000.

[35] Y. P. Fallah, P. Nasiopoulos, and R. Sengupta, “Fair scheduling for real-time multimedia support in IEEE 802.16 wireless access networks,” in *IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM’10)*, June 2010, pp. 1–9.

[36] Y. P. Fallah and H. Alnuweiri, “Analysis of temporal and throughput fair scheduling in multirate WLANs,” *Computer Networks*, vol. 52, Nov. 2008.

[37] Y. Fang, I. Chlamtac, and Y.-B. Lin, “Modeling PCS networks under general call holding time and cell residence time distributions,” *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 893–906, Dec. 1997.

[38] ——, “Call performance for a PCS network,” *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 8, pp. 1568–1581, Oct. 1997.

[39] Y. Fang and I. Chlamtac, “Teletraffic analysis and mobility modeling of PCS networks,” *IEEE Transactions on Communications*, vol. 47, no. 7, pp. 1062–1072, July 1999.

[40] A. Feldmann, “Impact of non-Poisson arrival sequences for call admission algorithms with and without delay,” in *IEEE Global Telecommunications Conference (GLOBECOM’96)*, vol. 1, Nov. 1996, pp. 617–622.

[41] G. Foschini, B. Gopinath, and Z. Miljanic, "Channel cost of mobility," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 414–424, Nov. 1993.

[42] M. Ghaderi and R. Boutaba, "Call admission control in mobile cellular networks: a comprehensive survey," *Wireless Communications and Mobile Computing*, vol. 6, no. 1.

[43] A. Goldsmith, *Wireless Communications.* Cambridge University Press, 2005.

[44] D. Grillo, R. Skoog, S. Chia, and K. Leung, "Teletraffic engineering for mobile personal communications in itu-t work: the need to match practice and theory," *IEEE Personal Communications*, vol. 5, no. 6, pp. 38–58, Dec. 1998.

[45] R. Guerin, "Channel occupancy time distribution in a cellular radio system," *IEEE Transactions on Vehicular Technology*, vol. 36, no. 3, pp. 89–99, Aug. 1987.

[46] G. Harine, R. Marie, R. Puigjaner, and K. Trivedi, "Loss formulas and their application to optimization for cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 3, pp. 664–673, May 2001.

[47] H. Hidaka, K. Saitoh, N. Shinagawa, and T. Kobayashi, "Teletraffic characteristics of cellular communication for different types of vehicle motion," *IEICE Transactions on Communications*, vol. E84-B, no. 3, pp. 558–565, March 2001.

[48] ——, "Vehicle motion in large and small cities and teletraffic characterization in cellular communication systems," *IEICE Transactions on Communications*, vol. E84-B, no. 4, pp. 805–813, April 2001.

[49] ——, "Self-similarity in cell dwell time caused by terminal motion and its effects on teletraffic of cellular communication networks," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Science*, vol. E85-A, no. 7, pp. 1445–1453, sep 2002.

[50] D. Hong and S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Transactions on Vehicular Technology*, vol. 35, no. 3, pp. 77–92, Aug- 1986.

[51] http://www.omnetpp.org.

[52] Y.-H. Hwang, S.-K. Noh, W.-J. Seok, and S.-H. Kim, "An effective resource management for fair call admission control using vertical handoff in 4g wireless networks," in *IEEE Vehicular Technology Conference (VTC-Spring 2008)*, May 2008, pp. 2188–2192.

[53] M. Ivanovich, M. Zukerman, P. Fitzpatrick, and M. Gitlits, "Performance between circuit allocation schemes for half- and full-rate connections in GSM," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 7, pp. 423–440, July 1992.

[54] V. Iversen, *Teletraffic Engineering and Network Planning.*

[55] C. Jedrzycki and V. Leung, "Probability distribution of channel holding time in cellular telephony systems," in *IEEE 46th Vehicular Technology Conference (VTC'97*, vol. 1, Apr-May 1996, pp. 247–251.

[56] H. Jiang and S. Rappaport, "Hand-off analysis for CBWL schemes in cellular communications," in *Third Annual International Conference on Universal Personal Communications*, Sep.-Oct. 1994, pp. 496–500.

[57] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey," *IEEE Personal Communications*, vol. 3, no. 3, pp. 10–31, Jun. 1996.

[58] F. Khan and D. Zeghlache, "Effect of cell residence time distribution on the performance of cellular mobile networks," in *IEEE 47th Vehicular Technology Conference (VTC'97)*, vol. 2, May 1997, pp. 949–953.

[59] ——, "Performance analysis of link adaptation in wireless personal communication systems," in *IEEE International Conference on Communications (ICC'97)*, vol. 3, Jun. 1997, pp. 1287–1291.

[60] K.-I. Kim and S.-H. Kim, “A light call admission control with inter-cell and inter-service fairness in heterogeneous packet radio networks,” *IEICE Transactions on Communications*, vol. E88-B, no. 10, pp. 4064–4073, Oct. 2005.

[61] L. Kleinrock, *Queueing systems. Volume I: Theory.* John Wiley&Sons, Inc., 1975.

[62] E. Kwon, J. Lee, K. Jung, and S. Ryu, “A Performance Model for Admission Control in IEEE 802.16,” in *Wired/Wireless Internet Communications (WWIC’05)*, ser. Lecture Notes in Computer Science, 2005.

[63] Y. H.-L. J. Kwon, E. and K. Jung, “Markov model for admission control in the wireless AMC networks,” *IEICE Trans. Commun.*, vol. E89-B, no. 8, pp. 2230–2233, Aug. 2006.

[64] W. Li and A. Alfa, “A PCS network with correlated arrival process and splitted-rating channels,” *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1318–1325, Jul. 1999.

[65] Q. Liu, S. Zhou, and G. Giannakis, “Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1142–1153, May 2005.

[66] J. G. Markoulidakis, J. E. Dermitzakis, G. L. Lyberopoulos, and M. E. Theologou, “Optimal system capacity in handover prioritised schemes in cellular mobile telecommunication systems,” *Computer Communications*, vol. 23, no. 5-6, pp. 462–475, 2000.

[67] I. Martin-Escalona, F. Barcelo, and J. Casademont, “Teletraffic simulation of cellular networks: modeling the handoff arrivals and the handoff delay,” in *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC’02)*, vol. 5, Sept. 2002, pp. 2209–2213.

[68] R. Murillo-Perez, C. Rodriguez-Estrello, and F. Cruz-Perez, “The impact of mobility on ofdma-based cellular systems with reuse partitioning,” in *IEEE Global Telecommunications Conference (GLOBECOM’09)*, Dec. 2009, pp. 1–7.

[69] P. V. Orlik and S. Rappaport, “On the handoff arrival process in cellular communications,” *Wireless Networks*, vol. 7, pp. 147–157, 2001.

[70] P. Orlik and S. Rappaport, “Traffic performance and mobility modeling of cellular communications with mixed platforms and highly variable mobilities,” in *IEEE 47th Vehicular Technology Conference (VTC)*, vol. 2, May 1997, pp. 587–591.

[71] ——, “A model for teletraffic performance and channel holding time characterization in wireless cellular communication with general session and dwell time distributions,” *IEEE Journal on Selected Areas in Communications,*, vol. 16, no. 5, pp. 788–803, jun 1998.

[72] S. Pattaramalai, V. Aalo, and G. Efthymoglou, “Call completion probability with weibull distributed call holding time and cell dwell time,” in *IEEE Global Telecommunications Conference (GLOBECOM’07)*, Nov. 2007, pp. 2634–2638.

[73] ——, “Evaluation of call performance in cellular networks with generalized cell dwell time and call-holding time distributions in the presence of channel fading,” *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 3002–3013, Jul. 2009.

[74] V. Paxson and S. Floyd, “Wide area traffic: the failure of poisson modeling,” *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, Jun. 1995.

[75] V. Pla and V. Casares-Giner, “Effect of the handoff area sojourn time distribution on the performance of cellular networks,” in *4th International Workshop on Mobile and Wireless Communications Network*, 2002, pp. 401–405.

[76] M. Rajaratnam and F. Takawira, "Hand-off traffic modelling in cellular networks," in *IEEE Global Telecommunications Conference (GLOBECOM'97)*, vol. 1, Nov. 1997, pp. 131–137.

[77] ——, "Nonclassical traffic modeling and performance analysis of cellular mobile networks with and without channel reservation," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 3, pp. 817–834, May 2000.

[78] ——, "Handoff traffic characterization in cellular networks under nonclassical arrivals and service time distributions," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 4, pp. 954–970, Jul. 2001.

[79] S. Rappaport, "Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms," vol. 140, no. 5, Oct. 1993, pp. 389–401.

[80] C. Rodriguez-Estrello, F. Cruz-Perez, and L. Ortigoza-Guerrero, "Residence times relationships in wireless communication systems with differentiated quality zones," in *IEEE 60th Vehicular Technology Conference (VTC-Fall 2004*, vol. 5, Sept. 2004, pp. 3414–3418.

[81] ——, "Performance Evaluation of CDMA Cellular Systems Considering both the Soft Capacity Constraint and Users' Smooth Random Mobility," in *IEEE International Conference on Communications (ICC'07)*, Jun. 2007, pp. 4098–4103.

[82] C. Rodriguez-Estrello, G. Hernandez-Valdez, and F. Cruz-Perez, "System-level analysis of mobile cellular networks considering link unreliability," *IEEE Transactions on Vehicular Technology,*, vol. 58, no. 2, pp. 926–940, Feb. 2009.

[83] S. M. Ross, *Introduction to probability models*, 9th ed. ELSEVIER, 2007.

[84] M. Ruggieri, F. Graziosi, and F. Santucci, "Modeling of the handover dwell time in cellular mobile communications systems," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 2, pp. 489–498, May 1998.

[85] T. S. Rappaport, *Wireless Communications: Principles and Practice.* Prentice Hall, Inc., 2002.

[86] K. Saitoh, H. Hidaka, N. Shinagawa, and T. Kobayashi, “Vehicle motion in large and small cities and teletraffic characterization in cellular communication systems,” *IEICE Transactions on Communications*, vol. E82-B, no. 12, pp. 2055–2060, Dec. 1999.

[87] M. Schwartz, *Mobile Wireless Communications.* Cambridge University Press, 2005.

[88] A. Sgora and D. Vergados, “Handoff prioritization and decision schemes in wireless cellular networks: a survey,” *IEEE Communications Surveys Tutorials*, vol. 11, no. 4, pp. 57–77, quarter 2009.

[89] M. Sidi and D. Starobinski, “New call blocking versus handoff blocking in cellular networks,” *Wireless Networks*, vol. 3, no. 1, pp. 15–27, March 1997.

[90] A. Simonsson and T. Lundborg, “Large scale mobility measurement in two metropolitan networks,” in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08)*, Sept. 2008, pp. 1–5.

[91] C. So-In, R. Jain, and A.-K. Al-Tamimi, “A scheduler for unsolicited grant service (UGS) in IEEE 802.16e mobile WiMAX networks,” *IEEE Systems Journal,*, vol. 4, no. 4, pp. 487–494, 2010.

[92] C. So-In, R. Jain, and A.-K. A. Tamimi, “Capacity evaluation for ieee 802.16e mobile wimax,” *Journal of Computer Systems, Networks, and Communications*, vol. 1, no. 1, pp. 1–12, January 2010.

[93] ——, “Deficit round robin with fragmentation scheduling to achieve generalized weighted fairness for resource allocation in IEEE 802.16e mobile WiMAX networks,” *Journal on Future Internet*, vol. 2, no. 4, pp. 446–468, Oct. 2010.

[94] ——, “Generalized weighted fairness and its application for resource allocation in IEEE 802.16e mobile WiMAX,” in *The 2nd International Conference on Computer and Automation Engineering (ICCAE’10)*, vol. 1, Feb. 2010, pp. 784–788.

[95] B. Soong and J. Barria, “A Coxian model for channel holding time distribution for teletraffic mobility modeling,” *Communications Letters, IEEE*, vol. 4, no. 12, pp. 402–404, Dec. 2000.

[96] A. Spedalieri, I. Martin-Escalona, and F. Barcelo, “Simulation of teletraffic variables in umts networks: impact of lognormal distributed call duration,” in *IEEE Wireless Communications and Networking Conference (WCNC’05)*, vol. 4, Mar. 2005, pp. 2381–2386.

[97] S.-L. Su, J.-Y. Chen, and J.-H. Huang, “Performance analysis of soft handoff in CDMA cellular networks,” *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 9, pp. 1762–1769, Dec. 1996.

[98] C. Tarhini and T. Chahed, “System capacity in OFDMA-based WiMAX,” in *International Conference on Systems and Networks Communications (ICSNC’06)*, Oct. 2006.

[99] ——, “AMC-aware QoS proposal for OFDMA-based IEEE 802.16 WiMAX systems,” in *IEEE Global Telecommunications Conference (GLOBECOM’07)*, Nov. 2007, pp. 4780–4784.

[100] ——, “On capacity of OFDMA-based IEEE802.16 WiMAX including adaptive modulation and coding (AMC) and inter-cell interference,” in *Local Metropolitan Area Networks, 2007. LANMAN 2007. 15th IEEE Workshop on*, 2007, pp. 139–144.

[101] ——, “Density-based admission control in IEEE802.16e Mobile WiMAX,” in *1st IFIP Wireless Days (WD’08)*, 24-27 2008, pp. 1–5.

[102] ——, “On mobility of voice-like and data traffic in IEEE 802.16e,” in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, Nov. 2008, pp. 1–5.

[103] ——, “QoS-oriented resource allocation for streaming flows in IEEE 802.16e mobile WiMAX,” *Telecommun. Syst.*, vol. 46, no. 1, pp. 1–7, Jan. 2010.

[104] S. Tekinay and B. Jabbari, “Handover and channel assignment in mobile cellular networks,” *IEEE Communications Magazine*, vol. 29, no. 11, pp. 42 –46, nov 1991.

[105] E. van Doorn and A. Ta, “Proofs for some conjectures of Rajaratnam and Takawira on the peakedness of handoff traffic,” *IEEE Transactions on Vehicular Technology*, vol. 52, no. 4, pp. 953–957, Jul. 2003.

[106] N. Vassileva, Y. Koycheryavy, and F. Barceló-Arroyo, “Guard capacity implementation in OPNET modeler WiMAX suite,” in *International Conference on Ultra Modern Telecommunications, ICUMT09*, Oct. 2009, pp. 1–6.

[107] H. Wang and V. Iversen, “Teletraffic performance analysis of multiclass OFDM-TDMA systems with AMC,” in *Wireless Systems and Mobility in Next Generation Internet*, ser. Lecture Notes in Computer Science, Springer Berlin/Heidelberg.

[108] X. Wang and P. Fan, “Channel holding time in wireless cellular communications with general distributed session time and dwell time,” *IEEE Communications Letters*, vol. 11, no. 2, pp. 158–160, Feb. 2007.

[109] R. J. WiMAX Forum, “WiMAX System Evaluation Methodology v2.1,” pp. 1–209, Jul. 2008.

[110] A. Xhafa and O. Tonguz, “Does mixed lognormal channel holding time affect the handover performance of guard channel scheme?” in *IEEE Global Telecommunications Conference (GLOBECOM’03)*, vol. 6, Dec. 2003, pp. 3452–3456.

[111] ——, “Handover performance of priority schemes in cellular networks,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 1, pp. 565–577, Jan. 2008.

[112] E. Yavuz and V. Leung, “Modeling channel occupancy times for voice traffic in cellular networks,” in *IEEE International Conference on Communications (ICC’07)*, Jun. 2007, pp. 332–337.

[113] H. Zeng and I. Chlamtac, “Handoff traffic distribution in cellular networks,” in *IEEE Wireless Communications and Networking Conference (WCNC’99)*, 1999, pp. 413–417.

[114] M. Zhang and C.-T. Lea, “Impact of mobility on CDMA call admission control,” *IEEE Transactions on Vehicular Technology,*, vol. 55, no. 6, pp. 1908–1922, Nov. 2006.

[115] Y. Zhang, “Handoff performance in wireless mobile networks with unreliable fading channel,” *IEEE Transactions on Mobile Computing,*, vol. 9, no. 2, pp. 188–200, Feb. 2010.

[116] Y. Zhang and B.-H. Soong, “Performance of mobile networks with wireless channel unreliability and resource insufficiency,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 5, pp. 990–995, May 2006.

[117] Y. Zhang, S. Xiao, M. Zhou, and M. Fujise, “Resource occupancy time in wireless networks,” in *IEEE International Conference on Communications (ICC’06)*, vol. 10, Jun. 2006, pp. 4440–4444.

[118] M. Zonoozi and P. Dassanayake, “User mobility modeling and characterization of mobility patterns,” *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 7, pp. 1239–1252, Sep. 1997.

# Appendixes

# List of acronyms

| | |
|---|---|
| AC | Admission control |
| AMC | Adaptive Modulation and Coding |
| ARQ | Automatic Repeat Request |
| BC | Busy Channels |
| BS | Base Station |
| CDMA | Code Division Multiple Access |
| CHT | Channel Holding Time |
| CPU | Central Processing Unit |
| CV | Coefficient of Variation |
| DCA | Dynamic Channel Allocation |
| DL | Downlink |
| EV-DO | Evolution-Data Optimised |
| FCA | Fixed Channel Allocation |
| FIFO | First In First Out |
| FUSC | Full Usage of Subchannels |
| GCQ | Guard Channel with Queueing |
| GCS | Guard Channel Scheme |
| GSM | Global System for Mobile communications |

| | |
|---|---|
| HARQ | Hybrid Automatic Repeat Request |
| HE2b | Hyper-Exponential with 2 Balanced stages distribution |
| HCA | Hybrid Channel Allocation |
| HSDA | High-Speed Data Access |
| IEEE | Institute of Electrical and Electronics Engineering |
| IMT-Advanced | International Mobile Telecommunications systems (IMT)–Advanced development |
| ITU-R | International Telecommunication Union–Radiocommunication Standartization Sector |
| ITU-T | International Telecommunication Union–Telecommunication Standartization Sector |
| LTE | Long Term Evolution |
| MAN | Metropolitan Area Network |
| MAP | Markov Arrival Process |
| MCS | Modulation and Coding Scheme |
| MIMO | Multiple In Multiple Out |
| MMPP | Markov Modulated Poisson Process |
| MRT | Mean Remaining Time scheme |
| MS | Mobile Station |
| MU | Mobile User |
| NPS | Non-Priority (also Non-Prioritisation) Scheme |
| OFDM | Orthogonal Frequency Division Multiplexing |
| OFDMA | Orthogonal Frequency Division Multiple Access |

| | |
|---|---|
| PAMR | Public Access Mobile Radio |
| PASTA | Poisson Arrivals See Time Averages |
| PCT-I | Pure Chance Traffic type I |
| PHY | Physical layer |
| PUSC | Partial Usage of Subchannels |
| SIR | Signal to Interference Ratio |
| SNR | Signal to Noise Ratio |
| QAM | Quadrature Amplitude Modulation |
| QoS | Quality of Service |
| QPSK | Quadrature Phase Shift Keying |
| SOHYP | Sum Of HYyper-Exponential distribution |
| UGS | Unsolicited Grant Service |
| UL | Uplink |
| UMTS | Universal Mobile Telecommunications System |
| WiMAX | Worldwide interoperability for Microwave Access |
| WCDMA | Wideband Code Division Multiple Access |
| xG | x-th Generation |

# List of Principal Symbols

| | |
|---|---|
| $A$ | offered traffic load |
| $A_c$ | carried traffic load |
| $\alpha$ | mobility factor that shows the average number of handoffs that a call goes through assuming there are infinite resources (no handoff dropping) |
| $C$ | capacity, that is the total number of system resources (channels or slots) |
| $C_g$ | number of guard slots |
| $E[X]$ | mean |
| $1/\eta$ | mean of the cell residence time |
| $1/\eta_i$ | mean of the cell residence time in zone $i$ |
| $\hat{f}(t\|\epsilon)$ | conditional probability density function of the residual lifetime |
| $g$ | guard channels |
| $\bar{h}(\epsilon)$ | expected value of the residual lifetime provided that $X = \epsilon$ age was attained |
| $\lambda$ | mean of the aggregate call arrival rate |
| $\lambda_n$ | mean of the new call arrival rate |
| $\lambda_n^i$ | mean of the new call arrival rate in zone $i$ |
| $\lambda_h$ | mean of the handoff call arrival rate |

| | |
|---|---|
| $m_1$ | mean |
| $1/\mu$ | mean of the unencumbered call duration |
| $1/\mu_r$ | mean of the channel holding time |
| $MRT$ | Mean Remaining Time metric |
| $n_i$ | number of active users (ongoing calls) in zone $i$ |
| $N_s^n$ | number of served new calls |
| $N_b^n$ | number of blocked new calls |
| $N_s^h$ | number of served handoff calls calls |
| $N_s^d$ | number of dropped handoff calls |
| $p_i$ | the area of zone $i$ expressed as a fraction of the total cell area |
| $Pb$ | Probability of blocking a new call |
| $Pd$ | Probability of dropping an ongoing call |
| $Pft$ | Probability of forced call termination |
| $q_{i+1,i}$ | probability that a mobile user in an outer zone $i+1$ will move to a neighbour zone $i$ that is closer to the base station |
| $s_i$ | slot capacity in zone $i$ |
| $\sigma$ | Standard deviation of a random variable |
| $TT$ | Time Threshold |
| $Var[X]$ | Variance |

# Publications

The research work presented in Part I was done at the Department of Telematics Engineering, Technical University of Catalonia (UPC), Spain, whereas the research presented in Part II was carried out at the Department of Communications Engineering, Tampere University of Technology (TUT), Finland. The thesis research work was carried out primarily by the author therefore, the thesis author is the main contributing author of publications P[1]-P[6] and P[8]. Naturally, the work was guided by the supervisors and the discussions between the author and Assoc. Prof. Francisco Barceló and Prof. Yevgeni Koucheryavy have shaped the investigations and their reporting as well as the obtained results. The main results from Part II are to be published. Publication P[7] is an outcome of the international collaboration between the two institutions—UPC and TUT—in which the concepts and ideas originally proposed by the thesis author and reported in Part II were extended to the more general case of elastic traffic in addition to voice traffic. The thesis author has participated in discussions that defined and led to the development of the extended model presented in P[7]. The thesis author has not participated in the script nor article writing.

## 8.1 Publications related to the thesis topic

[P1] N. Vassileva, and F. Barceló-Arroyo, "A new CAC policy based on traffic characterization in cellular networks," in *Proceedings of the Wired/Wireless Internet Conference*, WWIC'08 – LNCS 5031, Springer, pp. 1-12, May 2008, Tampere (FINLAND).

[P2] N. Vassileva and F. Barcelo-Arroyo, "Performance of a traffic-based handover method in high-mobility scenarios," in *Proc. 14th IEEE*

*International Conference on Parallel and Distributed Systems*, ICPADS'08, pp. 873-878, Dec. 2008, Melbourne (AUSTRALIA).

[P3] N. Vassileva and F. Barcelo-Arroyo, "Evaluation of CAC methods under non-Poisson arrival traffic in cellular networks," in *Proc. 11th IEEE International Conference on Communication Systems*, ICCS'08, pp. 919-925, Nov. 2008, Guangzhou (CHINA).

[P4] N. Vassileva and F. Barcelo-Arroyo, "Validation of a handover priority method using statistical traffic profile estimates in the context of general holding time distributions," in *Proc. 18th IEEE International Conference on Computer Communications and Networks*, ICCCN'09, pp. 1-7, Aug. 2009, San Francisco (USA).

[P5] N. Vassileva, Y. Koycheryavy, and Barceló-Arroyo, "Admission control for supporting active communication sessions in mobile WiMAX networks," in *Proc. 3rd ERCIM Workshop on eMobility* held in conjunction with WWIC'09, pp. 88-89, May 2009, Enschede (THE NETHERLANDS).

[P6] N. Vassileva, Y. Koycheryavy, and Barceló-Arroyo, "Guard capacity implementation in OPNET modeler WiMAX suite," in *Proc. International Conference on Ultra Modern Telecommunications*, ICUMT09, pp. 1-6, Oct. 2009, St. Petersburg (RUSSIA).

[P7] T. Efimushkina, N. Vassileva, D. Moltchanov, and Y. Koucheryavy, "Analytical Performance Evaluation of a WiMAX Cell with VoIP/Elastic Data Traffic," in *Proc. 11th IEEE Consumer Communications & Networking Conference*, IEEE CCNC'11, pp. 1-6, Jan. 2011, Las Vegas (USA).

[P8] N. Vassileva and F. Barceló-Arroyo, "Nueva poltica de control de admisin basada en caracterizacin de trfico en redes celulares," in *Proc. Jitel'08*, p. 1-7, Sept. 2008, Alcal de Henares (SPAIN).