Load Balancing in Heterogeneous Networks with LTE

Pedro Manuel Marques Ganço

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Examination Committee

Chairperson: Prof. José Eduardo Charters Ribeiro da Cunha Sanguino

Supervisor: Prof. Luís Manuel de Jesus Sousa Correia

Members of the Committee: Prof. António José Castelo Branco Rodrigues

February 2015

To my parents, for everything

Acknowledgements

I would like to start by thanking my thesis supervisor, Prof. Luís M. Correia. It has been an amazing opportunity to work under his supervision. Not only for his always present support and great advice on how to do a master thesis, but mainly for his knowledge on work methods and life in general that made him one of the best professors I had the pleasure to find at Instituto Superior Técnico. Without him, this thesis would not be a reality and I want to thank him for the chance of letting me "grow".

A great part of this thesis was done at GROW (Group for Research On Wireless) within an enthusiastic telecommunication work group. It has been great to share ideas with my fellow M.Sc. Students and other members of the group, namely David Doutor with whom I "balanced the load" of developing the simulator for this thesis, and Jorge Venes, a former GROW member who gave us few but decisive inputs to proceed our quest.

To my good old friends and to the good new ones I have come to find at IST, I want to thank you all for making my life worthwhile. A special thank you note goes to Nuno Graça, for his friendship and encouragement during all the ups and downs of the last few years and also to my friends of Setúbal, Pedro Lourenço, Gonçalo Oliveira, Paulo Pita and Miguel Silva.

I have been fortunate enough to study at IST alongside with Tiago Carreira, Pedro Venâncio, Pedro Diogo, Henrique Mendes, and then join NEEC (Núcleo de Estudantes de Engenharia e Computadores) where I met lots of good friends worth keeping. I am proud of all the great things we have accomplished and the difference we felt we were making at IST. A final word of appreciation goes to all the members of *Confraria*, for all the enthusiasm, hangouts and friendship.

Last but not least, I want to thank my parents and sisters for the unconditional love and support at home. None of this would have been possible without my parents, who always kept pushing me forward. I own them everything I have and the person I have become today.

Abstract

LTE networks are being deployed in urban scenarios, in several frequency bands, superimposing cells with GSM and UMTS ones. Thus, there is a structure of heterogeneous cells, not only within LTE itself, but also with the other systems. The scope of the thesis is to find a method for balancing the load among the mentioned systems; one defines metrics for load computation based on the number of fixed radio resources each system has, and then analyses in a time-based network simulator a model that moves traffic to balance the load. One also studied the impact of varying the user density, radio configuration, number of base stations and other relevant scenarios, and then designed a scenario that enables the performance analysis of the model developed for load balance. The model allowed to increase the performance in heterogeneous scenarios in overloading situations. The main results show some improvement in the overall QoS parameters, while increasing the average load for all RATs and increasing the average number of users connected per second. The main results in terms of QoS are the reduction of drop rate calls from 2.3% to 1.2% and the reduction of blocking probability from 15.2% to 3.4% measured for the same scenario. The model developed for this work may be used as a general framework for load balancing in heterogeneous networks.

Keywords

Heterogeneous Networks, Load Balancing, Radio Resource Management, LTE

Resumo

As redes LTE estão a ser implementadas nas grandes cidades, em várias bandas de frequências, com as estações base co-localizadas com as existentes em GSM e UMTS, criando uma estrutura de células heterogéneas, não só com LTE, mas também com os outros sistemas. O âmbito desta tese foi o de desenvolver um método de distribuição da carga dos diferentes sistemas, e para tal, definiram-se métricas de cálculo de carga com base no número fixo de recursos rádio de cada sistema, para seguidamente analisar num simulador temporal de rede um modelo que mova tráfego para balancear carga. Estudou-se o impacto de variar a densidade de utilizadores, configuração rádio, número de estações base e outros cenários relevantes, onde seguidamente é criado um cenário capaz de testar a análise de desempenho do modelo de distribuição de carga. O modelo permite aumentar o desempenho da rede heterogénea em ambientes de sobrecarga. Os resultados principais em termos de qualidade de serviço verificaram melhorias e ao mesmo tempo aumentam a carga média dos sistemas e número de utilizadores servidos. Os resultados principais de qualidade de serviço consistem na redução da queda de chamadas de 2.3% para 1.2% e a redução de probabilidade de bloqueio de 15.2% para 3.4% para um mesmo cenário simulado. O modelo desenvolvido para este trabalho pode ser usado como uma base para a distribuição de carga em redes heterogéneas.

Palavras-chave

Redes Heterogéneas, Distribuição de carga, Gestão de Recursos Rádio, LTE.

Table of Contents

List of Figures

List of Tables

List of Acronyms

1G	First Generation
2G	Second Generation
3G	Third Generation
4G	Fourth Generation
3GPP	3rd Generation Partnership Project
ARFCN	Absolute Radio Frequency Channel Number
BCH	Broadcast Channel
BE	Best Effort
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CAC	Call Admission Control
CAGR	Compound Annual Growth Rate
CBR	Constant Bit Rate
CCCH	Common Control Channel
CDMA	Code Division Multiple Access
CEPT	European Conference of Postal and Telecommunications Administration
CN	Core Network
CS	Circuit Switch
DCCH	Dedicated Control Channel
DL	Downlink
E-UTRAN	Evolved UTRAN
EDGE	Enhanced Data rates for GSM Evolution
eNB	Evolved Node B
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FHO	Forced HO
GGSN	Gateway GPRS Support Node
GMSC	Gateway MSC
GMSK	Gaussian Minimum Shift Keying
GPRS	General Packet Radio Service

GROW	Group for Research On Wireless
GSM	Global System for Mobile Communications
HHO	Horizontal Handover
HLR	Home Location Register
HN	Heterogeneous Network
HNP	Heterogeneous Network Performance
HO	Handover
HSDPA	High-Speed Downlink Packet Access
HSPA	High-Speed Packet Access
HSUPA	High-Speed Uplink Packet Access
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardisation Sector
JD-TCDMA	Joint Detection Time/Code Division Multiple Access
LoS	Line of Sight
LTE	Long Term Evolution
M2M	Machine-to-Machine
ME	Mobile Equipment
MIMO	Multiple-Input Multiple-Output
MME	Mobility Management Entity
MS	Mobile Station
MSC	Mobile Switching Centre
NB	Node B
NLoS	Non Line of Sight
NMT	Nordic Mobile Telephone
NSS	Network and Switching Subsystem
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OVSF	Orthogonal Variable Spreading Factor
OSS	Operation Support Subsystem
PCRF	Policy and Charging Resource Function
P-GW	PDN Gateway
PDN	Packet Data Network
PS	Packet Switch
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature-Phase Shift Keying
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block

REF	Reference Scenario
RF	Radio Frequency
RNC	Radio Network Controller
RRH	Remote Radio Head
RRM	Radio Resource Management
RRU	Radio Resource Unit
S-GW	Serving Gateway
SAE	System Architecture Evolution
SC-FDMA	Single Carrier FDMA
SF	Spreading Factor
SGSN	Serving GPRS Support Node
SISO	Single Input Single Output
SMS	Short Message Service
SNR	Signal to Noise Ratio
SON	Self-Organising Networks
TACS	Total Access Communication System
TCH	Traffic Channel
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TRX	Transceivers
TS	Timeslot
TTI	Transmission Time Interval
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
USIM	Universal Subscriber Identity Module
UTRAN	UMTS Terrestrial RAN
VHO	Vertical Handover
VLR	Visitor Location Register
WCDMA	Wideband CDMA

List of Symbols

$\mathrm{a_{pd}}$	Average power decay
B_G	Average global bitrate
B_{RAN}	Average bitrate per RAN
B_{srv}	Bitrate per service
B	Total bandwidth available
$\bar{B}_u$	Average bandwidth of the RBs allocated per user
d	Cell radius
D_r	Drop rate
E_b	Average bit energy
E_b/N_0	Signal to Noise Ratio (SNR)
$G_{G_{multi-user\ div}}$	Capacity gain obtained due to multi-user diversity
G_{pos}	Capacity gain obtained due to users positioning in the cell
G_r	Gain of the receiving antenna
G_t	Gain of the transmitting antenna
h_b	Height of the BS antenna
H_B	Height of the buildings
h_m	Height of the Mobile Terminal
HNP^{cum_s}	Partial HNP cumulative mean at simulation run s
HNP^{cum_S}	Total HNP cumulative mean
$\overline{L_p}$	Average path loss between Node B and UE
M_{code}	Modulation and coding rate
N_0	Noise spectral density
$\bar{N}_{BSMT}$	Average number of BSs reachable per MT
N_c	Number of incoming voice and video calls

N_{ca}	Number of carriers
N_{cb}	Number of blocked voice and video calls
N_{CH}	Number of available channels
N_{co}	Number of codes
N_{fc}	Number of radio channels available for traffic
N_{fd}	Number of frames being delayed
N_{HHOa}	Number of Horizontal HO attempts
N_{HHOf}	Number of Horizontal HO failures
N_{RB}	Number of RB
N_{RRUa}	Number of RRU available for traffic
N_{RRUt}	Total number of RRUs
N_{s}	Total number of sessions
N_{sd}	Number of dropped sessions
N_{sym}	Number of symbols transmitted per time unit
N_{TCH}	Number of traffic channels
N_{TRX}	Number of TRX (Transceivers) placed in one GSM site
N_{TS}	Number of users allocated into TSs during a TTI per frequency channel
N_{usr}	Number of users connected to a BS
$N_{U_{GSM}}$	Number of GSM users at full rate in a cell
$N_{U_{LTE}}$	Number of LTE users in a cell
$N_{U_{UMTS}}$	Maximum number of UMTS users in a cell
N_{VHOa}	Number of Vertical HO attempts
N_{VHOf}	Number of Vertical HO failures
P_b	Blocking probability
P_{HHOf}	Probability of Horizontal HO failure
P_{VHOf}	Probability of Vertical HO failure
P_{TU}	Power assigned for a specific user
w_B	Building separation
w_s	Street width

α	Code orthogonality of user
ϕ	Incidence angle
η_{sch}	Scheduler efficiency
ρ	Average load of a given RAN
ρ_{thr}	Load threshold
$\bar{\tau}$	Average delay

List of Software

CRRM SimCell	Time-based Radio Resources Simulator
MS Excel 2013	Calculation tool and tables processor
MS Visual Studio 2005	Integrated development tool
MS Word 2013	Word processor
SmartDraw 7	Flowcharts drawing tool

Chapter 1

Introduction

This chapter briefly presents a chronological evolution of some wireless telecommunications systems before introducing the motivation, objectives and scope of the thesis.

At the end of the chapter there is also a detailed description of the structure of the thesis.

1.1 Overview and Motivation

Since Marconi's first invention of the wireless telegraphic system in 1894, the world has seen a tremendous evolution in the field of cordless telecommunications.

The first generation (1G) for mobile telecommunications started to be deployed by many independent systems in the early 1980s, such as Total Access Communication System (TACS) or the Nordic Mobile Telephone (NMT) for Europe. This 1G technologies relied on analogic communications before being deprecated against the inevitable digital systems that were about to come [SeTB11].

In 1982 the European Conference of Postal and Telecommunications Administration (CEPT) – predecessor of ETSI (European Telecommunications Standards Institute) – started the development of a new 2G system enabling new services, such as mobile roaming and SMS. In a matter of years, this 2G system would become a worldwide success, the Global System for Mobile Communications (GSM) [Moli11].

Just in the turn of the new millennium, in 2002, the Universal Mobile Telecommunications System (UMTS) was introduced to the market, with higher data rates enabling new services and supporting the ever growing data demand.

The fourth generation (4G) of mobile communications was developed to supply those data demands, with downlink data rates surpassing hundreds of Mbps, making use of new architecture and modulation schemes in the radio interface. The Long Term Evolution Advanced (LTE-A) is the most recent generation deployed, but the telecom world is already thinking about next generation requirements and forecasts what future services and data demands will be.

Figure 1-1 gives a perspective on the number of mobile broadband subscribers and market penetration around the world.

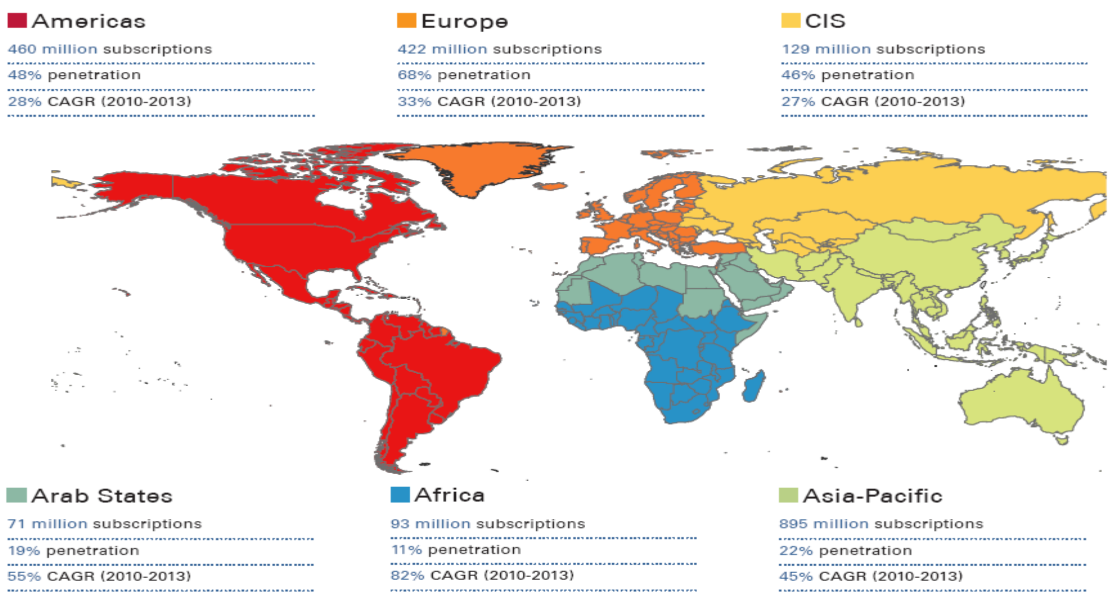


Figure 1-1 - Mobile broadband penetration (extracted from [ICT13]).

In Figure 1-2, one can also see that the number of mobile devices using 2G will most probably decrease, but will still be very important in many years to come when facing the evolution of data,

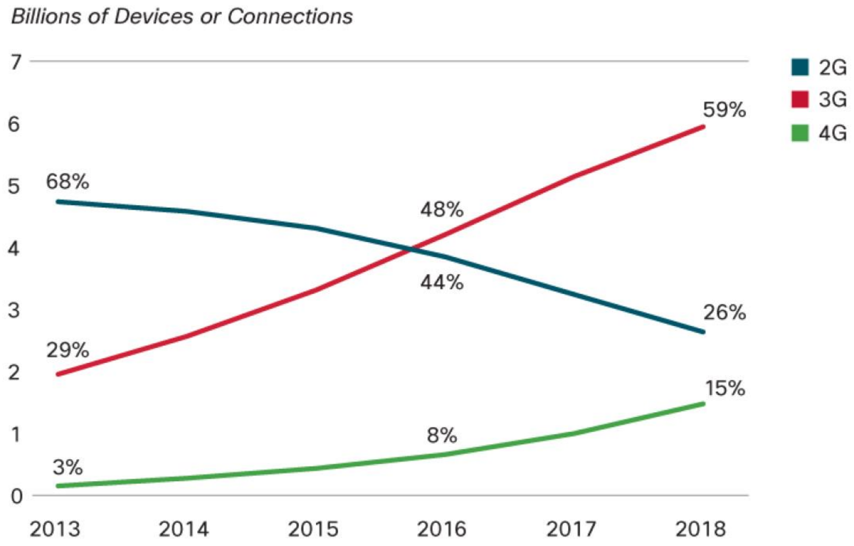


Figure 1-2 - Global Mobile Devices and Connections (extracted from [CISC13]).

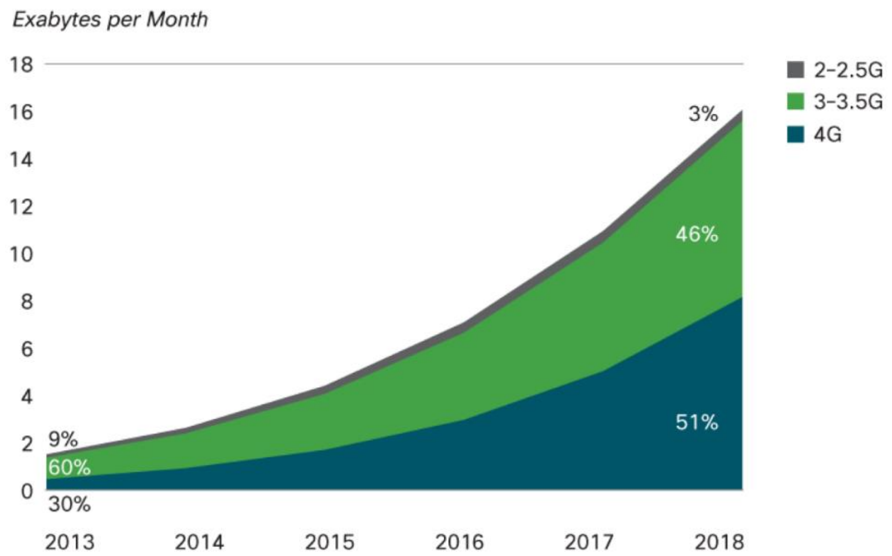


Figure 1-3 - 2G, 3G and 4G Penetration trends (extracted from [CISC13]).

3G and 4G systems will reinforce their market penetration, hence, it is crucial for mobile operators to balance the load among systems, and efficiently use their network resources. According to [CISC13], the trend to offload traffic among different systems will increase, as shown in Figure 1-4. This is due to the increasing mobile capabilities for connection simultaneously to more than one radio technology.

In face of the presented exploding traffic challenges and heterogeneous environment, where many different systems co-exist, it is the operator's best interest to be able to balance the load among existing systems.

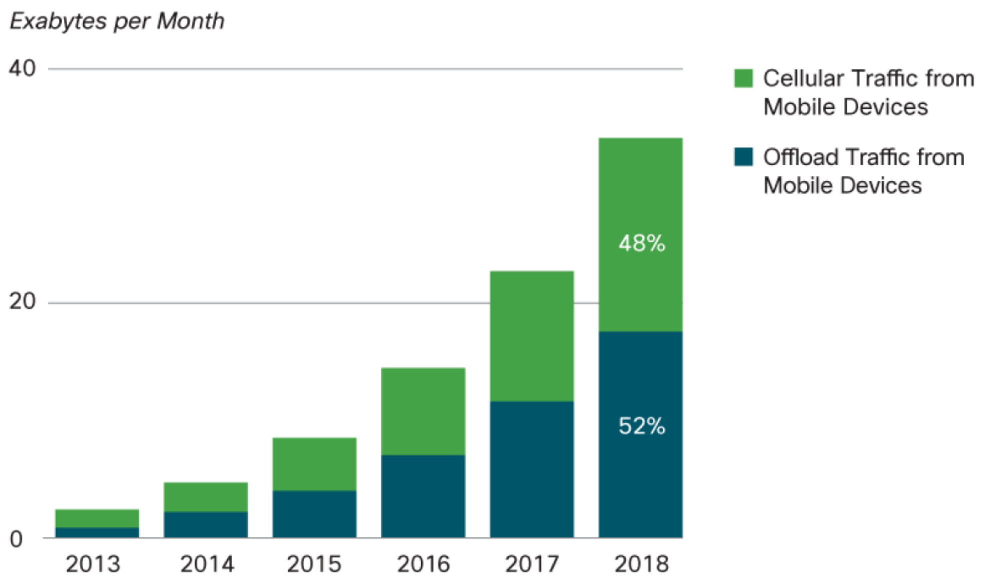


Figure 1-4 - Offload traffic trend (extracted from [CISC13]).

The scope of this thesis is to provide a model that enables load balancing among GSM, UMTS and LTE, taking quality of service into account. One focuses only on the mentioned three systems: GSM (2G), UMTS (3G) and LTE (4G), since they represent a huge portion of the total mobile users among the 3 generations deployed around the world, the available literature shows commons methods for offloading traffic among heterogeneous systems.

The current work makes use of a simulation tool developed in a PhD thesis by Serrador [Serr12] with the Visual Studio 2005 platform to implement a complex CRRM (Common Radio Resources Management) simulator that generates users, traffic and a general support system to handle most of the common radio functionalities of a real wireless communication environment. In its first version, the simulator had modules for UMTS and WiFi only, but it was later on improved in an M.Sc. Thesis by Venes [Vene09] to add WiMAX capabilities. For the purposes of the present thesis on load balancing in heterogeneous networks with LTE, one has added GSM and LTE functional modules, as well as new M2M (machine-to-machine) services. The simulator, like its previous version, is only considering the downlink channel.

The development of the simulator modules for GSM and LTE comprises a vast amount of working hours, since after implementation it had to be optimised, tested and assessed before testing the models to obtain results with statistical significance.

This thesis aims to bring some insights on ways to measure load in a heterogeneous radio environment with GSM, UMTS and LTE, as well as to provide some practical ways of sharing load among these different RATs for the purpose of service quality improvement.

1.2 Structure of the Dissertation

The present chapter makes a brief overview of the mobile wireless communication's history evolution, showing the motivation behind the thesis. It also states the scope of the developed studies, an introduction to the challenges ahead, and the author's contributions and expectations for the study on load balancing in heterogeneous networks.

Chapter 2 contains the essential aspects of GSM, UMTS and LTE, in order to provide a perspective on the systems' differences, and to subsequently find a common way to measure load, taking all systems' peculiarities into account. The chapter ends with a section on load balance and its state of the art.

The models developed in this thesis can be found in Chapter 3, where expressions for load metrics are explained in detail, and methodologies on how to tackle the load balancing challenge are shown. This chapter also describes the software tool used to test and infer the models designed.

In Chapter 4, the scenarios to test the model and results drawn from such models are presented.

The conclusions of this work can be found in Chapter 5. This short chapter makes an overview of the thesis theme, models and main results obtained in simulation. It also enunciates a paragraph on future work suggestions in order to improve and push forward the work on load balance.

Chapter 2

Basic Aspects

This chapter provides an overview of the GSM, UMTS and LTE systems and presents the state of the art regarding load balancing.

2.1 GSM

This section reviews the main aspects of GSM regarding its networks architecture and radio interface. Most of the concepts briefly introduced here are explained in further detail in [Moli11] and [Walk02].

2.1.1 Network Architecture

Figure 2-1 shows the GSM network architecture, splitting the Packet Switch (PS) from the Circuit Switch (CS) elements in the Core Network (CN). The PS elements, such as the SGSN (Serving GPRS Support Node) and the GGSN (Gateway GPRS Support Node), add packet transmission capabilities, and were only included in latter stages of GSM evolution, which relied only on CS in its genesis. The network architecture of GSM can be divided into three main parts: the Base Station Subsystem (BSS), the Network and Switching Subsystem (NSS) and the Operation Support Subsystem (OSS).

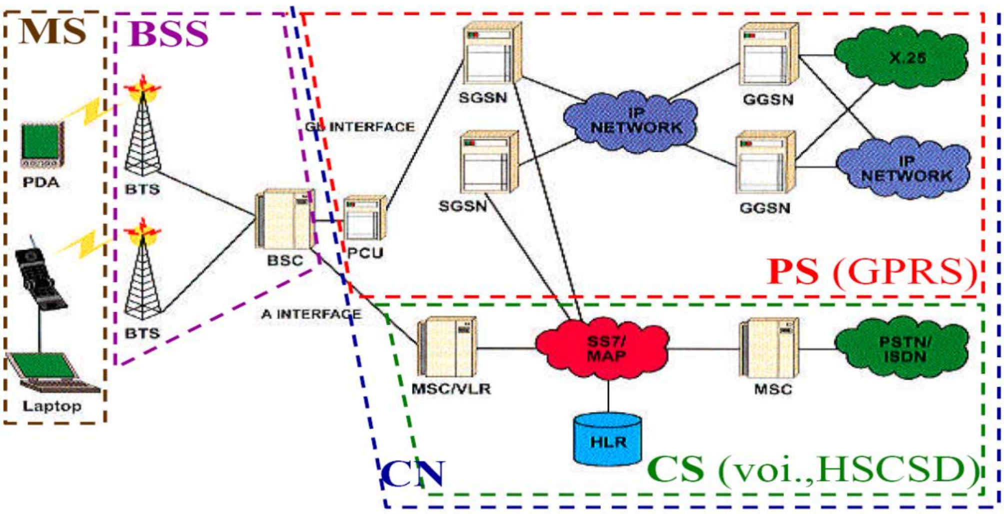


Figure 2-1 - GSM architecture (extracted from [Corr13]).

The BSS is composed of two main elements: the Base Transceiver Station (BTS) that takes care of the radio link between all Mobile Stations (MS) and the network within its coverage area, and the Base Station Controller (BSC), that is in charge of a set of BTSs simultaneously and responsible for mainly control functions, such as handover. The main functionalities of the BSS are the maintenance of link quality, taking care of channel assignment, handover, power control, coding and encryption. The NSS consists essentially of the Mobile Switching Centre (MSC) and several databases. The MSC manages handovers when an MS is moving to a cell belonging to a different BSC and also routes traffic to external networks. In order to grant user mobility, there is a set of databases to store user’s location information, i.e., the Home Location Register (HLR) and the Visitor Location Register (VLR). Finally, the OSS is responsible for managing the entire network. It controls the accounting services of each subscriber, performs maintenance tasks to check the network elements functionality, and collects data regarding quality of links and traffic demands.

2.1.2 Radio Interface

In GSM, Frequency Division Duplex (FDD) is used, splitting Uplink (UL) and Downlink (DL) traffics into separate bands with a 45 MHz frequency spacing between the UL and DL. Both UL and DL frequency ranges, shown in Table 2-1, are partitioned into a 200 kHz grid and consecutively numbered – Absolute Radio Frequency Channel Number (ARFCN) –, for each of these 200 kHz sub-bands, 8 different users are allocated, splitting the time axis into timeslots of 576.92 µs.

Table 2-1 - Frequency ranges for UL and DL in GSM [Moli11].

	Uplink [MHz]	Downlink [MHz]
GSM 900	[890, 915]	[935, 960]
GSM 1800	[1710, 1785]	[1805, 1880]

GSM uses a combination of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA), where a physical channel is no more than a combination of a specific timeslot index and an ARFCN, each of these channels having a bandwidth and a frequency separation of 200 kHz.

The physical channel transmits both user data and signalling traffic, depending on the logical channel associated with it. Many logical channels are used for different purposes, such as:

- Traffic Channel (TCH): used to transmit payload data consisting, e.g., encoded voice data;
- Broadcast Channel (BCH): only for DL, carrying the information necessary for an MS to synchronise in both time and frequency, and to establish a connection to the BS;
- Common Control Channel (CCCH): used for a BS to transmit information to all MSs, or for the initial setup stage of a new connection, since there is no dedicated channel between the BS and the new MS;
- Dedicated Control Channel (DCCH): transmits the required signalling information during a connection, being a bidirectional channel dedicated to one specific connection.

Each type of data is transmitted in specific timeslots, thus, it is necessary to map the logical channel onto the physical ones. In order to better understand the mapping between logical and physical channels, one must first introduce the time structure of physical channels. The basic period, a frame, consists of 8 timeslots. Each frame has a duration of 4.61 ms, and 26 frames are combined to make a multiframe. A superframe has a duration of 6.12 s and consists of 51 multiframes. The time structure for traffic channels is used for the mapping between logical and physical channels. For instance, not all timeslots are used for TCH. For every periodic multiframe, 24 frames are used for TCH and the remaining 2 frames are used for signalling through DCCHs. More detailed information on this topic can be found on [Moli11].

GSM networks were first designed mainly for voice communications. The first step to transfer data in GSM was done by HSCSD (High Speed Circuit Switched Data), which allowed data transmissions over CS, with a maximum of 4 time slots per connection reaching up to 57.6 kbps data rates.

Then came GPRS, introducing PS communications. In GPRS, four coding schemes (CS1-CS4) are considered, depending on the radio interface coding and the number of time slots to be used. It allows up to 8 time slots per connection, which combining with a coding scheme CS4 provides a maximum theoretical bitrate of 171.2 kbps.EDGE is based on a new modulation in the radio interface. Instead of using only the usual Gaussian Minimum Shift Keying (GMSK), it also uses Phase Shift Keying (PSK), achieving higher data rates per time slot. It has 9 possible Modulation Coding Schemes (MCS1-MCS9), modulated in GMSK from MCS1 to MSC4 and modulated with 8-PSK from MCS5 till MSC9. Given all these different approaches, data rates naturally vary depending on the services used, Table 2-2.

Table 2-2 - Data rates in GSM (adapted from [Corr13] and [LeMa01]).

<table>
<tr><th colspan="2">System and Service</th><th>Data rate per Timeslot [kbps]</th><th>n, number of time slots per connection</th><th>Maximum data rate [kbps]</th></tr>
<tr><td colspan="2">Voice</td><td>22.8</td><td>1</td><td>22.8</td></tr>
<tr><td colspan="2">Data</td><td>9.6</td><td>1</td><td>9.6</td></tr>
<tr><td colspan="2">HSCSD (data)</td><td>14.4</td><td>1,2,3,4</td><td>57.6</td></tr>
<tr><td rowspan="4">GPRS (data)</td><td>CS1</td><td>9.05</td><td rowspan="13">1,2,…,8</td><td>72.4</td></tr>
<tr><td>CS2</td><td>13.4</td><td>107.2</td></tr>
<tr><td>CS3</td><td>15.6</td><td>124.8</td></tr>
<tr><td>CS4</td><td>21.4</td><td>171.2</td></tr>
<tr><td rowspan="9">EDGE (data)</td><td>MCS1</td><td>8.8</td><td>70.4</td></tr>
<tr><td>MCS2</td><td>11.2</td><td>89.6</td></tr>
<tr><td>MCS3</td><td>14.8</td><td>118.4</td></tr>
<tr><td>MCS4</td><td>17.6</td><td>140.8</td></tr>
<tr><td>MCS5</td><td>22.4</td><td>179.2</td></tr>
<tr><td>MCS6</td><td>29.6</td><td>236.8</td></tr>
<tr><td>MCS7</td><td>43.8</td><td>350.4</td></tr>
<tr><td>MCS8</td><td>54.4</td><td>435.2</td></tr>
<tr><td>MCS9</td><td>59.2</td><td>473.6</td></tr>
</table>

2.2 UMTS

This section presents an overview of UMTS, regarding its network architecture and radio interface based on [HoTo07] and [Moli11].

2.2.1 Network Architecture

The UMTS network architecture was built upon GSM's, and kept unchanged throughout 3G evolution from Release 99 to Release 7, being also divided into 3 main parts: User Equipment (UE), UMTS Terrestrial Radio Access Network (UTRAN) and Core Network (CN), as represented in Figure 2-2.

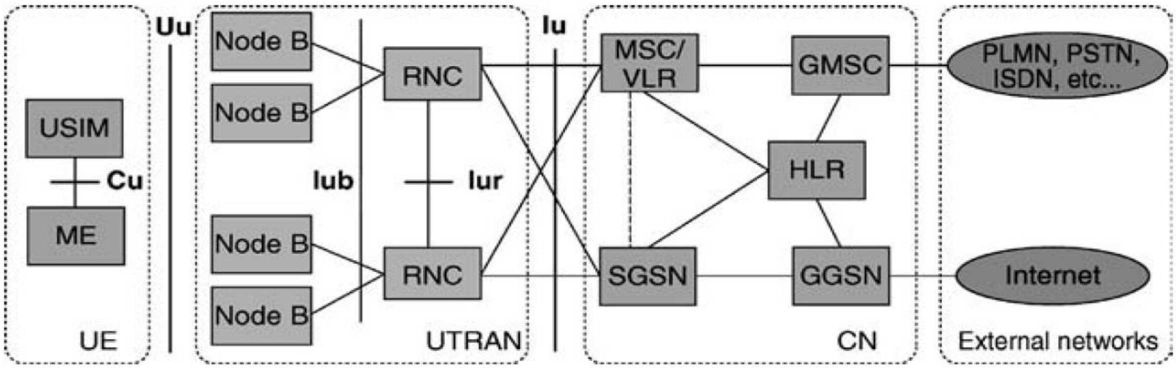


Figure 2-2 - UMTS architecture (extracted from [HoTo04]).

The User Equipment consists of the Mobile Equipment (ME), which is the radio terminal that communicates with UTRAN over the Uu interface, and the Universal Subscriber Identity Module (USIM), which contains the subscriber identity and other user-related information.

The UTRAN consists of two elements:

- the Node B, which is responsible for the connection of the UE with the network and radio resource management functions;
- and the Radio Network Controller (RNC), connecting the Node Bs to the CN and controlling the radio resources of its base stations. RNCs may share traffic and signalling data, a feature not available in GSM, where the communication between two BSCs has to go through the CN.

The Core Network, besides having the usual storage of user's information in databases (such as HLR), routes the traffic into CS or PS networks. The main elements responsible for PS are SGSN and GGSN, whereas for CS, it is the Mobile Switching Centre (MSC) and Gateway MSC (GMSC) responsibility to route the incoming traffic to CS networks.

2.2.2 Radio Interface

FDD is the relevant mode of operation in UMTS, and its frequency bands are as shown in Table 2-3, with 60 MHz of bandwidth for both UL and DL, and a frequency spacing of 5 MHz for each radio channel.

Table 2-3 - Frequency ranges for UL and DL in UMTS [Moli11].

	Uplink [MHz]	Downlink [MHz]
UMTS FDD	[1920, 1980]	[2110, 2170]

UMTS employs Code Division Multiple Access (CDMA) to distinguish different users and sources in the air interface. This technique consists of making user data go through a process of channelisation and scrambling. The channelisation process spreads the original data, by multiplying them with quasi-random bits (called chips) derived from CDMA spreading codes. A chip rate of 3.84 Mcps leads to a carrier bandwidth of approximately 4.4 MHz, hence, the output signal is now spread over a wide bandwidth where each user has a unique orthogonal code so they can operate in the same frequency band simultaneously without causing harmful interference. Channelisation codes are therefore used to separate the physical data from control channels in the same terminal in UL, whereas for the DL they are used for the separation of DL connections to different users within one cell. Spreading codes are based on the Orthogonal Variable Spreading Factor (OVSF) technique, allowing the Spreading Factor (SF) to change while maintaining orthogonality between different spreading codes of different lengths. After channelisation, the scrambling code is used on top of spreading to distinguish terminals in the UL and for the separation of cell sectors in the DL, and does not affect either the symbol rate or the bandwidth coming out of channelisation. One important feature is that it supports highly variable data rates for different users in a cell, allowing changes of data capacity every 10 ms (which is the initial frame duration), due to the assignment of variable SFs to MTs. Further enhancements to 3G were made to enable higher data rates, such as the appearance of High Speed Packet Access (HSPA) for DL and UL. In HSDPA the SF is fixed to 16, allowing a maximum theoretical data rate of 1.4 Mbps per code, where using more efficient modulations and an Adaptive Modulation and Coding (AMC) scheme according to the link quality achieved higher peak data rates than previous releases. The user data rates may vary on many factors such as the link quality, the service and Release, although generally, one may obtain the theoretical data rates in Table 2-4.

Table 2-4 - Data rates in UMTS (extracted from [Corr13]).

Service	**Release**	**Data rate [kbps]**	
		Uplink	**Downlink**
Voice	99	12.2	12.2
Data	99	< 64	< 384
	5 (HSDPA)	< 384	< 14 400
	6 (HSUPA)	< 5 800	< 14 400
	7 (HSPA+)	< 11 500	< 28 000

2.3 LTE

This section provides some basic aspects on LTE, namely about its architecture and radio interface, being based on [SeTB11] and [Moli11].

2.3.1 Network Architecture

The LTE network architecture represented in Figure 2-3 has less elements than 2G and 3G ones, and can be split into three main domains: the User Equipment (UE), Evolved UTRAN (E-UTRAN) and the Evolved Packet Core (EPC). The UE condenses the Universal Subscriber Identity Module (USIM), responsible for authentication and security procedures, and the Terminal Equipment (TE). The E-UTRAN bridges between the UE and the EPC, being simply composed of the evolved Node B (eNB). There is no need for a controller such as a BSC or RNC like in legacy 2G and 3G networks, respectively, since the eNB now performs all radio-related functionalities, being connected among each other via X2 interfaces. The EPC is responsible for the overall control of the UE and the establishment of the bearers and no longer supports legacy CS-based communications. Its main elements are the following:

- MME (Mobility Management Entity), is the main control element, responsible for functions such as mobility management, subscription profile management, service connectivity, authentication and security;
- HSS (Home Subscription Server), is the database server that stores the location and permanent user data;
- SAE-GW (System Architecture Evolution Gateway), combining the S-GW (Serving Gateway) responsible for the User Plane tunnel management and switching and the P-GW (Packet Data Network Gateway) responsible for routing the traffic to external packet data networks;
- PCRF (Policy and Charging Resource Function), responsible for policy and charging functions.

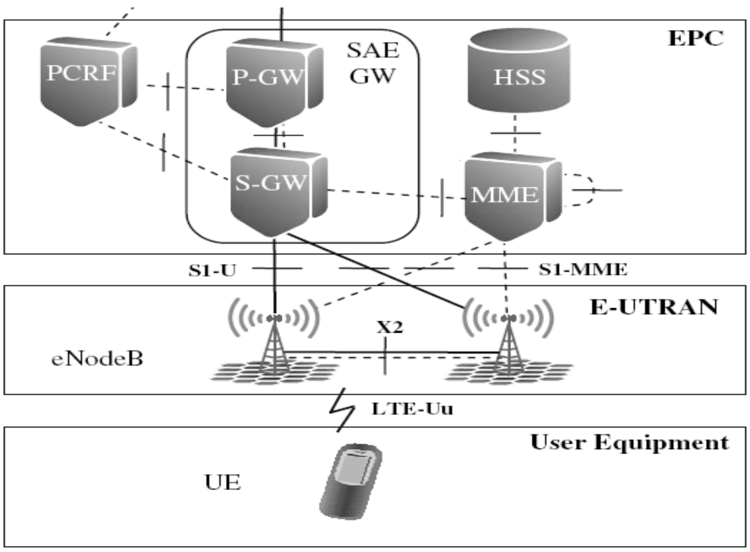


Figure 2-3 - LTE network architecture (adapted from [HoTo09]).

2.3.2 Radio Interface

LTE is typically spread along three bands of the spectrum (800 MHz, 1800 MHz and 2600 MHz), but this may differ slightly from country to country, thus, in Table 2-5 one presents the currently spectrum allocation for LTE FDD in Portugal [ANAC12].

Table 2-5 - Frequency ranges for UL and DL in LTE [ANAC12].

	Uplink [MHz]	Downlink [MHz]
LTE 800	[832, 862]	[791, 821]
LTE 1800	[1805, 1880]	[1710, 1785]
LTE 2600	[2630, 2690]	[2510, 2570]

The multiple access technique used in LTE differs for UL and DL: in DL one has Orthogonal Frequency Division Multiple Access (OFDMA) whereas for UL it is Single Carrier Frequency Division Multiple Access (SC-FDMA) in order to minimise power consumption at the UE domain.

A guard period, Cyclic Prefix (CP), is inserted between each transmitted symbol to prevent inter-symbol interference and reduce the equaliser's complexity. For a normal CP, each time slot of 0.5 ms has 7 OFDMA symbols or 6 OFDMA symbols in case of an extended CP. A set of 12 sub-carriers makes a Resource Block (RB), thus, occupying a 180 kHz bandwidth (12 sub-carriers $\times$ 15 kHz sub-carriers spacing) and having 84 Resource Elements (RE) (12 sub-carriers $\times$ 7 symbols, for normal CP).

The radio resources allocation in LTE is shown in Figure 2-4.

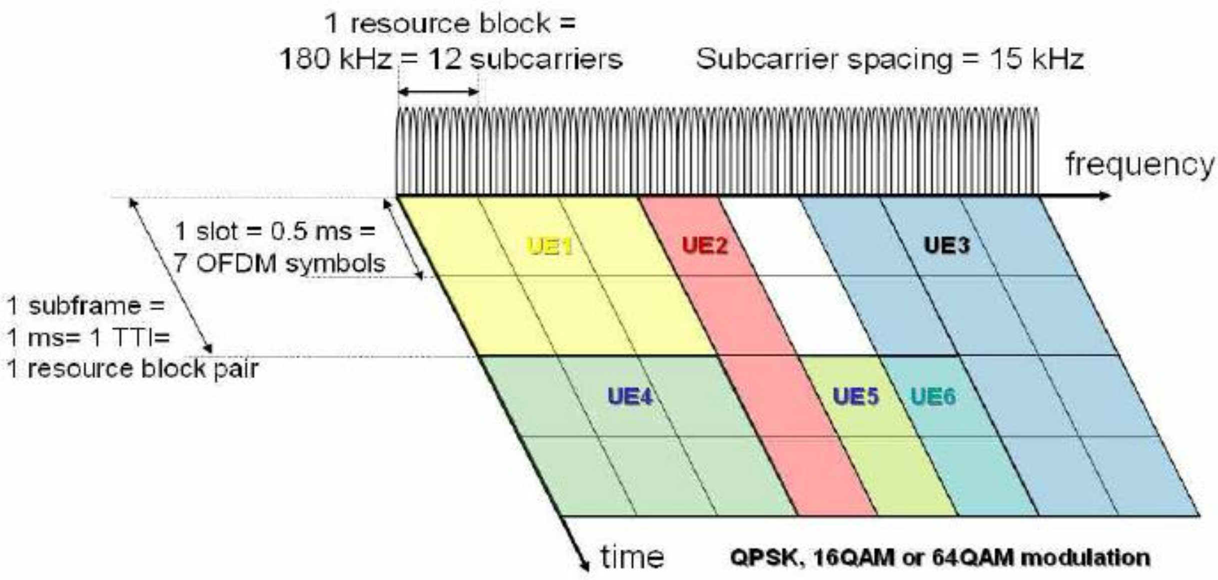


Figure 2-4 - Radio resources allocation in LTE in the time-frequency domain. (extracted from [Corr13]).

LTE allows up to six different bandwidths for the radio channels, as shown in Table 2-6, depending on the number of sub-carriers allocated in a period of time to a user.

Table 2-6 – Relation between the number of sub-carriers, RB and Bandwidth (adapted from [Corr13]).

Bandwidth [MHz]	1.4	3	5	10	15	20
Number of sub-carriers	72	180	300	600	900	1200
Number of Resource Blocks	6	15	25	50	75	100

Concerning modulation, LTE uses both Quadrature Phase Shift Keying (QPSK) and Quadrature Amplitude Modulation (QAM). For DL one has QPSK, 16QAM or 64QAM, whereas for UL, only UE of category 5, 7 or 8 allow a modulation up to 64QAM. Note that the first five categories are present in Release 8, 9 and 10, but categories 6, 7 and 8 were only introduced in Release 10; furthermore, not all UE categories support MIMO, which will restrain the peak throughput achievable by a UE, nonetheless, considering the maximum allowed modulation scheme and MIMO support, if available, one can obtain the peak throughput of UE per category in Table 2-7 for UL and DL.

Table 2-7 - UE's categories in LTE (adapted from [DaPS11]).

UE Category	**Peak throughput [Mbps]**	
	UL	**DL**
1	5	10
2	25	50
3	50	100
4	50	150
5	75	300
6	50	300
7	150	300
8	1500	3000

LTE's high peak data rates are achieved by using the maximum bandwidth of 20 MHz (in the first releases), 64QAM modulation and MIMO transmission. In Table 2-8 and

Table 2-9 one has the peak bit rate obtained for DL and UL, respectively, for a normal CP usage.

Regarding the present thesis, one focuses on the DL channel only and does not take the bitrate constraints imposed by MTs shown in Table 2-7, assuming that any user can accommodate any bitrate as service demands.

Table 2-8 - Downlink peak bit rates in LTE (adapted from [HoTo11]).

Downlink peak bit rates [Mbps]								
Resource blocks			Bandwidth [MHz]					
Modulation and coding	**Bits per symbol**	**MIMO**	**1.4**	**3**	**5**	**10**	**15**	**20**
QPSK ½	1	-	1.0	2.5	4.2	8.4	12.6	16.8
16QAM ½	2	-	2.0	5.0	8.4	16.8	25.2	33.6
16QAM ¾	3		3.0	7.6	12.6	25.2	37.8	50.4
64QAM ¾	4.5	-	4.5	11.3	18.9	37.8	56.7	75.6
64QAM 1/1	6	-	6.0	15.1	25.2	50.4	75.6	100.8
64QAM ¾	9	2 x 2	9.1	22.7	37.8	75.6	113.4	151.2
64QAM 1/1	12	2 x 2	12.1	30.2	50.4	100.8	151.2	201.6
64QAM 1/1	24	4 x 4	24.2	60.5	100.8	201.6	302.4	403.2

Table 2-9 - Uplink peak bit rates in LTE (adapted from [HoTo11]).

Uplink peak bit rates [Mbps]							
Resource blocks		Bandwidth [MHz]					
Modulation and coding	**Bits per symbol**	**1.4**	**3**	**5**	**10**	**15**	**20**
QPSK ½	1	1.0	2.5	4.2	8.4	12.6	16.8
16QAM ½	2	2.0	5.0	8.4	16.8	25.2	33.6
16QAM ¾	3	3.0	7.6	12.6	25.2	37.8	50.4
16QAM 1/1	4	4.0	10.1	16.8	33.6	50.4	67.2
64QAM ¾	4.5	4.5	11.3	18.9	37.8	56.7	75.6
64QAM 1/1	6	6.0	15.1	25.2	50.4	75.6	100.8

2.4 Load Balance

In this section, one presents some general ideas regarding the user capacity per system and vertical handover procedures necessary for load balancing, mainly based on [Saut11] and [HoTo09].

2.4.1 Capacity per System

Load metrics differ from system to system, and depend on the number of users in a given cell, services being provided, quality of service and many other different parameter. In Chapter 3, one tackles this issue in further detail, therefore, the following subsection only presents the computation of the maximum number of users in a given cell for the different systems studied in this thesis. In GMS, capacity depends mainly on the number RF Transceivers (TRX) installed in the BS, since each user is assigned its own timeslot (at full-rate). The number of radio channels is constant in GSM, each one carrying 8 timeslots, therefore, taking into consideration that 10% of the BS capacity is used for signalling and control purposes, the number of users at full-rate in one cell is given by [Silv12]:

$$N_{U_{GSM}} = (N_{TRX} \cdot 8) \cdot 0.9 \cdot \frac{N_{TCH}}{N_{CH}} \tag{2.1}$$

where:

- N_{TRX} : number of transceivers placed in one GSM site;
- N_{TCH} : number of traffic channels;
- N_{CH} : number of available channels.

As for UMTS, capacity strongly depends on the interference level among users, since all UEs are operating at the same frequency due to CDMA. Hence, for a specific data rate, the number of UMTS users in a cell can be approximated by [Silv12]:

$$N_{U_{UMTS}} = \frac{1 - 10^{0.1 \cdot \left(-P_{TU[\mathrm{dBW}]} - G_{t\,[\mathrm{dBi}]} - G_{r[\mathrm{dBi}]} + \left({}^{E_b}/_{N_0}\right)_{[\mathrm{dB}]} - G_{p[\mathrm{dB}]} + \overline{L_p}_{[\mathrm{dB}]} + N_{0[\mathrm{dBW}]} + 10\, a_{pd} \log\left(d_{[\mathrm{km}]}\right)\right)}}{\frac{{}^{E_b}/_{N_0}}{G_p}\ [(1-\alpha) + i]} \tag{2.2}$$

where:

- P_{TU} : power received by user, assumed to be the same for all users in site;
- G_t : gain of the transmitting antenna;
- G_r : gain of the receiving antenna;
- E_b : energy per bit;
- N_0 : noise power spectral density at the UE;
- G_p : processing gain of user, given by the ratio between the chip rate and the user bit rate;
- $\overline{L_p}$: average path loss between Node B and UE;
- a_{pd} : average power decay;
- d : cell radius;
- α : code orthogonality of user (typical values between 0.5 and 0.9);
- i : inter-cell interference ratio.

To conclude, the total number of users in an LTE cell, can be estimated by [Carr11]:

$$N_{U_{LTE}} = \frac{B_{[\mathrm{Hz}]}\ \eta_{sch}\ G_{pos}\ G_{G_{multi-user\ div}}}{\bar{B}_{u\ [\mathrm{Hz}]}} \tag{2.3}$$

where:

- B : total bandwidth available;
- G_{pos} : capacity gain obtained due to users positioning in the cell;
- $G_{G_{multi-user\ div}}$: capacity gain obtained due to multi-user diversity;
- $\bar{B}_u$: average bandwidth of the RBs allocated per user;
- η_{sch} : scheduler efficiency, chosen in [0,1].

2.4.2 Vertical Handover Procedures

GSM, UMTS and LTE are currently being explored by most operators simultaneously, and many MTs are equipped with antennas to work seamlessly with more than one system. Figure 2–5 shows the combined architecture that allows inter-working among GSM, UMTS and LTE. Since there are no direct interfaces between the different Radio Access Technologies (RATs), signalling is all carried via the EPC that needs to add new interfaces to connect to UTRAN and GERAN, namely interfaces S3, S4 and S12.

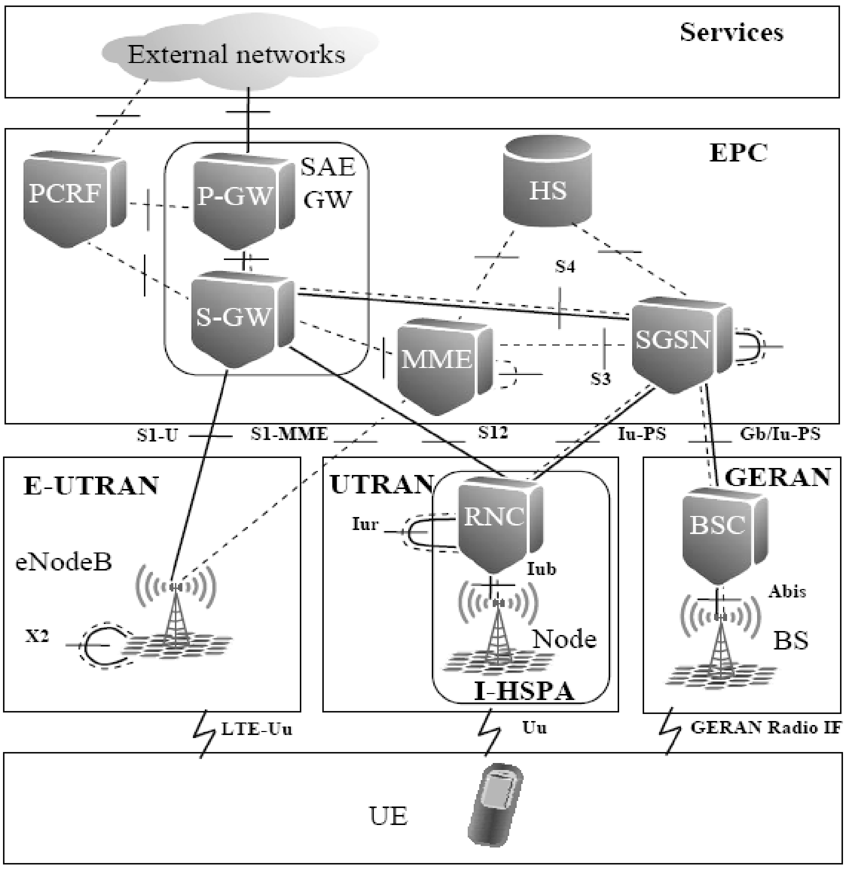


Figure 2-5 - Architecture with E-UTRAN and Legacy 3GPP RANs (adapted from [HoTo09]).

When UTRAN or GERAN is connected to the EPC, they may still operate as before, and for this purpose the S-GW assumes the role of the GGSN. The eNB must be able to coordinate UEs measuring UTRAN and GERAN cells, and perform HandOver (HO) decisions based on measurement results, thus E-UTRAN radio interface protocols have been appended to support this new feature. The intersystem HO steps follow one of the two following procedures depending on the source/target system:

- Handover from E-UTRAN to UTRAN or GERAN:
  The eNodeB requests the UE to measure the signal level of UTRAN or GERAN cells, and analyses the measurement reports. If the eNodeB decides to start the handover, it signals the need to the MME in the same way that it would signal inter-eNodeB HO when the X2 interface is not available. Thereafter, the eNodeB will receive the information needed for the HO command from the target access system via the MME. The eNodeB will send the HO command to the UE without the need for interpreting the content of this information.
- Handover from UTRAN or GERAN to E-UTRAN:
  The eNodeB allocates the requested resources, and prepares the information for HO command, which it sends to the MME, from where it is delivered to the UE through the other access system that originated the handover.

The inter-system HO is controlled by the source access system for starting the measurements and it is a backwards HO where the radio resources are reserved in the target system before the HO command is issued to the UE. Naturally, if a GERAN network does not support PS communications, the resources are not reserved before the HO. All the information from the target system is transported to the UE via the source system transparently, and the user data can be forwarded from the source to the target system to avoid loss of user data. In order to speed up the HO procedure there is no need for the UE to have any signalling to the core network and nowadays the interruption time in inter-system HOs can be as low as in the order of 50 ms. An overview of inter-system HO is shown in Figure 2-6, where the Serving GW can be used as the mobility anchor for the inter-system HO of the three access systems.

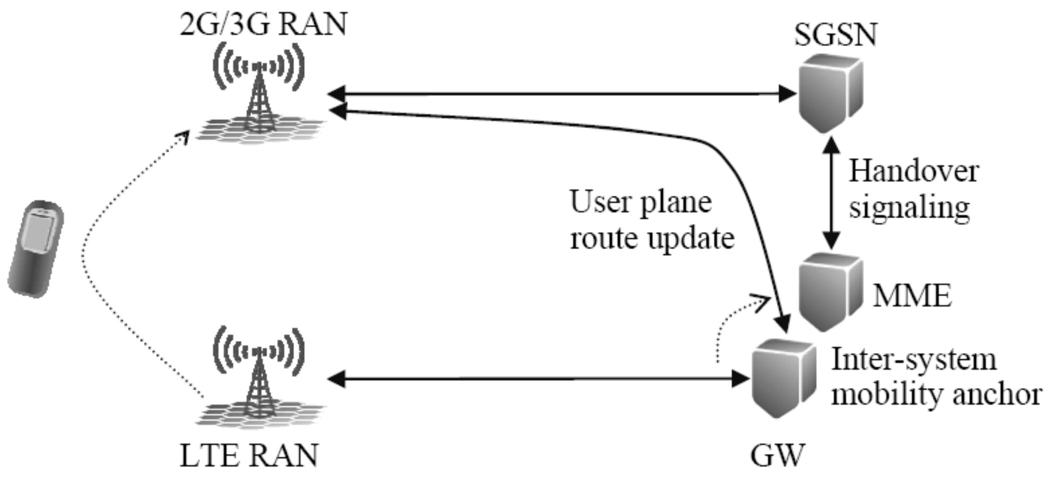


Figure 2-6 - Inter-system handover (extracted from [HoTo09]).

The HO between UMTS and GSM is as follows [Eric03]:

- Handover from UTRAN to GERAN

  When the signal strength falls below a given threshold, the UMTS network orders the MT to perform GSM measurements. When UTRAN receives the measurement report message sent from the MT, it initiates HO. UTRAN then asks the target BSS to reserve resources. The target BSS prepares a HO command message, which includes the details of the allocated resources. This GSM message, which is sent to the MT via the UMTS radio interface, is transferred within a container that is transparently passed on by the different network nodes. When the MT receives the HO command, it moves to the target GSM cell and establishes the radio connection in accordance with the parameters included in the HO command message. The MT indicates successful completion of HO by sending a HO complete message to the BSS, after which the GSM network initiates the release of the UMTS radio connection.

- Handover from GERAN to UTRAN

  The network orders the dual-mode MT to perform UMTS measurements by sending the measurement information message, which contains information on neighbouring UMTS cells and the criteria for performing and reporting measurements. When the criteria for HO to UMTS have been met, the BSS initiates the allocation of resources to the UMTS cell. Encapsulated in these messages, the BSS also sends information to UTRAN on the UMTS capabilities of the MT. When the resources of the UMTS target cell have been allocated, UTRAN compiles the HO-to-UTRAN-command message, which typically includes the identity of the pre-defined configuration for the service in use. This message is then sent transparently to the MT through the core network and BSS. When the MT receives the HO-to-UTRAN command message it tunes to the UMTS frequency and begins radio synchronisation. The MT then indicates that the HO was successful by sending the HO-to-UTRAN-complete message, after which the resources in GSM are released.

2.5 State of The Art

3GPP Release 8 introduced the concept of the Self-Organising Networks (SON) where performance parameters are adjusted dynamically and automatically. Up to this point, live networks were most of the times designed by over dimensioning network requirements to cover peak load situations, hence the need for new techniques of load balancing to efficiently use the scarce radio resources. The concept of Heterogeneous Networks in literature many times refer to overlapping cells from different sources within the same system, like a femto-cell inside a pico-cell whose load can be shared with each other. In this thesis, however, one considers a heterogeneous network to be a scenario with different Radio Access Technologies (RATs) coexist and are able to communicate, namely GERAN, UTRAN and E-UTRAN.

In order to balance the load in a heterogeneous network, one can employ many approaches, although, in the scope of this thesis, one focuses on balancing the load by periodic HOs of load and by Call Admission Control (CAC). The former is responsible by applying methodologies of Vertical Handover (VHO) or Horizontal Handover (HHO) between GSM, UMTS and LTE in order to harmonise the networks current load sharing, and the latter by the analysis for accepting or denying every new incoming call/session request and H of ongoing calls.

There are plenty of different ways suggested in literature to implement load balancing via VHO. In [FGSP05], Ferrús et al. suggest the use of parameters such as service type, network conditions, operators policies, user preferences and signal level for a VHO decision. In [CCOJ05], the authors suggest using a combination of signal level, service delay sensitivity, service financial cost, and mobile conditions to trigger VHO. Other authors, such as [ZdSc04] and [BiHJ03], give much importance to signal quality by triggering the VHO due to signal level, signal to interference ratio and BER (Bit Error Ratio) measurements. More focused on energy efficiency, in [KYII10] the selection of RAT is the one leading to lower energy consumption to increase MT battery life while respecting Quality-of-Service (QoS); when considering energy efficiency from the operators' viewpoint, in [EART11] a wide set of saving techniques are proposed to reduce networks infrastructure power consumption. In the current thesis, the triggering point for a HO in a heterogeneous networks is based on the idea that, once a given load threshold is reached, some of the traffic is handed over to a legacy system to free radio resources.

Chiu et al. [ChJa89] introduced for the first time a balance index to measure the balance of resources in a system, given by:

$$\xi_1 = \frac{(\sum_i \rho_i)^2}{n \sum_i \rho_i^2} \tag{2.4}$$

where:

- n : is the number of neighbouring BSs over which the load can be distributed;
- ρ_i : represents the load of BS i.

This balance index quantifies the balance among neighbouring BSs and equals 1 when all BSs have the same load and tends to $1/n$ when the load is severely unbalanced, therefore, its target is to be maximised. To define a BS load-state and define such a threshold parameter, δ, for a VHO decision, the definition of average load is given by:

$$\bar{\rho} = \frac{\sum_i \rho_i}{n} \tag{2.5}$$

where the BS load state is then:

- Under-loaded: when the BS load is below $\bar{\rho}$;
- Balanced: when the BS exceeds the load average by less than δ;
- Overloaded: when the BS load exceeds $\bar{\rho}$ by δ.

This so called "load" metric represents the occupation ratio of a BS, and can be calculated in different manners depending on the system, therefore, the load value of two different networks may not represent the same load situation. Some common methods for load computation are based on interference, [BeJu03], or throughput, [VeAl04], though, Nguyen-Vuong et al. [NgAg11] identified some limitations to this balance index and load metrics. A given user generates different added loads depending on the BS it is connected to. For example, given de scenario represented in Figure 2-7, where a new user within coverage of three BSs needs decide to which one to connect, the current load of BS A, BS B and BS C ($\rho_A = 0.8$, $\rho_B = 0.4$, and $\rho_C = 0.3$, respectively), according to the load index ξ_1 make the new user connect to BS C, although the added load of the new user make that same BS C be overloaded.

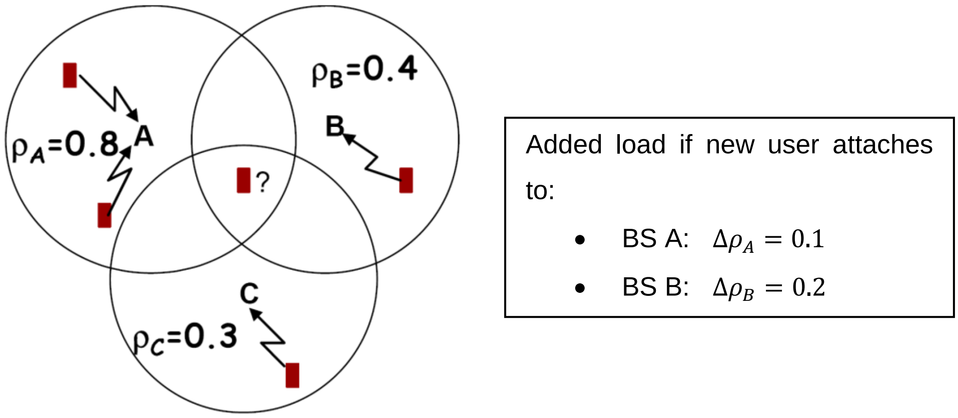


Figure 2-7 - Problem of using ξ_1 (adapted from [NgAg11]).

The best connection for the new user in this case would have been BS B. Concerning such limitations, [NgAg11] proposes a new algorithm to deal with LB in heterogeneous packet networks. First, a general load metric has been defined as the ratio of the required resources to the total ones, to hide the resources heterogeneity among different networks (see Chapter 3 for different systems load computation), thereafter, a new balance index was defined as:

$$\xi_2 = \sum_{i}^{n} max(\rho_i - \delta, 0) \tag{2.6}$$

The objective of the new LB algorithm is now to minimise ξ_2 until it is equal to zero or there is no further improvements to be made. The LB algorithm proposed by [NgAg11] is employed to prevent overload situations (by admitting or rejecting new communications) and employed also in managing ongoing communications (by forcing HO in imminent overload situations). A connection request to a specific BS is only accepted if the BS's load, including the contribution of the incoming communication, is below an admission threshold, otherwise, it is redirected to the least loaded overlapped access network.

The connection is rejected if there is no BS available for connection in the coverage area, unless it is a HO. Handovers are always accepted and this HO enforcement is done in a two-step process: in the first step, a move (I, J) is identified to move a mobile user M_i from an overloaded BS_0 to a suitable BS_J. If the network is still overloaded, the algorithm runs for a two-move operation moving a user M_i from BS_0 to BS_j and then move a user M_L from BS_j to BS_K until it can no longer improve the balance index

ξ_2. Figure 2-8 represents the algorithm explained above, where w_{ij} is the load contribution of user M_i at BS_j while M_i connects to BS_j.

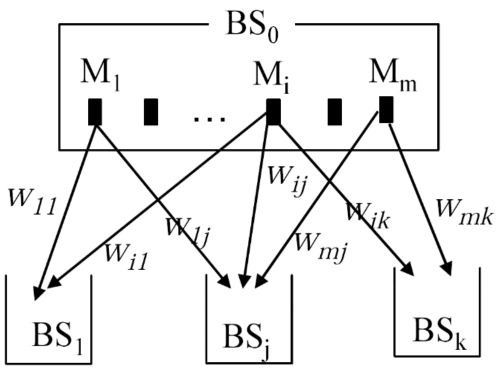


Figure 2-8 - LB algorithm scheme (extracted from [NgAg11]).

This thesis takes all the presented state of the art into consideration in Chapter 3, although one may outline the suggested algorithm of Nguyen-Vuong et al. [NgAg11] in the LB for VHO and HHO, whereas for the CAC, one based the ideas on [Serr12] and evaluated the performance of such developed models in Chapter 3 in heterogeneous environments containing all GSM, UMTS and LTE.

Chapter 3

Algorithms and Simulator Description

Chapter 3 contains the description of the models and algorithms developed for the load balancing in heterogeneous networks. Furthermore, this chapter also introduces the functionalities and characteristics of the simulator used to obtain the results.

3.1 Algorithms and Models

3.1.1 Propagation, Traffic Generation and User Mobility Models

The propagation model considered in this thesis is applied by taking urban scenarios as a reference, since it is in such scenarios that LTE is being mainly deployed, therefore, one considers the COST 231 Walfish-Ikegami Model, described in further detail in Annex A for the path loss calculation for all RATs. It is important to note that GSM operates at the 900 MHz band, UMTS at 2100 MHz and LTE at 2600 MHz, and since the validity range of the COST 231 Walfish-Ikegami Model goes from 800 MHz to 2000 MHz the implementation of the model introduces some propagation measurement errors. The buildings' height, streets' width and all the remaining parameters of the model are within their validity range and further characterised in the reference scenario on Chapter 4.

Regarding the models for traffic generation, one takes advantage of some of the work developed by Serrador [Serr12], who explains the models used for traffic generation for voice services, web browsing, video calls, e-mail, file sharing and music streaming services. On top of those, the present thesis contributes to the improvement of the existing simulator by adding new services, namely, Machine-to-Machine (M2M) communication, whose traffic models were adapted from the services already implemented, in order to add eHealth, smart meters, surveillance and domotics services. These new services were mapped onto the traffic models developed by [Serr12] according to their service classes and characteristics. One considers the use of the voice source model described in Annex B.1 for conversational Real Time (RT) voice calls and the model from Annex B.2 to generate both video calls and domotics services. Concerning the service classes for Non-Conversational Applications, such as file sharing, web browsing, e-mail, smart meters and e-Health, one applies the models from Annex B.3. For simplification purposes, streaming RT services, like music streaming and surveillance, are treated as being non-conversational, using just different parameters to approximate a real life traffic scenario. More details on the used traffic source models, and respective parameters, are addressed Annex B.

In order to reproduce MTs' behaviour, one implemented the Random Walk Mobility Model developed by [Serr12], which gives memory-less mobility patterns, as each step is calculated without any information from the previous one. At regular time intervals, both the direction (uniform distributed from 0º to 360º) and the speed of the MTs are updated. MTs' speed is combined with this mobility model, using a triangular distribution density model, being capable of generating different speeds for each MT, which is important for link load variation and HO generation due to propagation changes. Further information regarding mobility models can be found in Annex C

3.1.2 Load Metrics for Heterogeneous RATs

The great challenge in balancing the load among heterogeneous RATs is in finding a common load measurement for all RATs that take the system's differences in allocating resources into account. Consider for instance one overloaded UMTS BS transferring some load to a neighbouring LTE one.

The UMTS BS's release of MTs, via VHO, implying, e.g., a 20% reduction of its load would not necessarily imply an equivalent 20% increase on the LTE BS's load, since the radio resource allocation in UMTS and LTE differ, as seen in Chapter 2. Therefore, one must first define what load is and afterwards define a common metric to exchange load from users and services among BSs, regardless of their RAT.

The maximum number of users in a BS can be calculated according to the expressions in Section 2.4.1. This is a good starting point to obtain a threshold for the maximum number of users allocated to a given BS, but its value is hard to compare among different RATs, since it may depend on the interference levels, SNR and other highly variable (and hard to measure) parameters. Load could also be defined in terms of the BS's maximum throughput, power budget and many other ways, as shown in Section 2.5, but for the purpose of performing load balance in a heterogeneous RAN, one has taken the approach of finding a common metric for load focusing only on the radio resource allocation.

Henceforth, to compute the load of a BS, one defines a generic Radio Resource Unit (RRU) that represents the amount of radio resources a BS has to connect to MTs. The load of a BS, regardless of its RAT, can now be defined as follows:

$$\rho_{[\%]} = \left(1 - \frac{N_{RRUa}}{N_{RRUt}}\right) 100 \qquad (3.1)$$

where:

- N_{RRUa} : number of RRU available for traffic;
- N_{RRUt} : total number of RRU;

The concept of RRU hides the RATs differences and in order to quantify the total number of RRUs a given BS has, one must understand the fixed amount of radio resources any technology has to communicate and the way resources are consumed by users.

In a TDMA/FDMA system, like GSM, where one considers only voice, load consists mainly on the number of users one can accommodate for phone call services or SMS. The allocation of resources is done in periods of 4.61 ms of Timeslots (TS), where any given radio channel at full-rate transmission has capacity for 8 different users. GSM900's DL band ranges from 935 MHz to 960 MHz, therefore a separation of 200 kHz per channel gives us a theoretical maximum of 125 frequency channels, although, one should consider that more than one telecom operator uses the spectrum and that not all radio channels can be used just for traffic purposes.

A more realist approach to reality may consider the attribution of 10 radio channels for traffic purposes allocated to a single telecom operator, each carrying 8 TS, imposing a maximum of 80 radio resources per Time Transmission Interval (TTI) in a GSM BS, obtained by the following computation of total radio resources in GSM:

$$N_{RRUt_{GSM}} = N_{TS} \ N_{fc} \qquad (3.2)$$

where:

- N_{TS} : number of users allocated to TSs during a TTI in a radio channel (which equals to 8 for a full-rate transmission);
- N_{fc} : number of radio channels available for traffic.

The radio resources (timeslots) capacity can be taken from Table 2-2, where, for example, for voice calls, the data rate is 22.8 kbps.

Regarding UMTS's BSs, for the purpose of this thesis one will only consider the more recent radio access technology release, i.e., HSDPA. In this system, channel multiplexing is done in the time domain, where each TTI with 2 ms of duration can carry 480 symbols. Within each TTI, a maximum of 15 parallel codes per carrier can be assigned to one user (or shared among several) for traffic usage. This variable number of codes per carrier assigned to a user imposes the total number of RRU in UMTS:

$$N_{RRUt_{UMTS}} = N_{co} \quad N_{ca} \tag{3.3}$$

where:

- N_{co} : number of codes;
- N_{ca} : number carriers (typically, there are 3 carriers per BS).

Anytime a given MT is using a service in HSDPA, the number of RRUs this service consumes can be calculated as follows:

$$N_{RRUr_{UMTS}} = \left\lceil \frac{B_{srv\,[\mathrm{bps}]}}{N_{sym[\mathrm{symb/s}]} \quad M_{code\,[\mathrm{bits/symb}]}} \right\rceil \tag{3.4}$$

where:

- B_{srv} : service bitrate;
- N_{sym} : number of symbols transmitted in a second (480 symbols in a TTI, correspond to 240 000 symbols in a second);
- M_{code} : modulation and coding rate in use (e.g., a modulation of 16-QAM, and unitary coding rate, generates 4 bits per symbol).

In an OFDM system such as LTE, RBs are sent every 0.5 ms and its total number depends on the bandwidth as shown in Table 2-6, e.g., for a bandwidth of 20 MHz one has 100 RB. Note however that the TTI is 1 ms, hence, the available RBs per 1 ms is 200. The amount of RBs can be used as the measurement for the total number of RRU in LTE, $N_{RRUtLTE}$.

An LTE BS generates 7 OFDM symbols every 0.5 ms in the time-domain and 12 sub-carriers (1 RB) in the frequency domain, therefore, the required resources a given service demands in LTE is given by:

$$N_{RRUr_{LTE}} = \left\lceil \frac{B_{srv\,[\mathrm{bps}]}}{N_{RB} \quad N_{sym[\mathrm{symb/s}]} \quad M_{code\,[\mathrm{bits/symb}]}} \right\rceil \tag{3.5}$$

where:

- N_{RB} : number of RB;

Not all RRUs are used for traffic. Some are used for signalling, e.g., therefore to compute the number of RRU available for traffic, N_{RRUa}, one should consider that for UMTS, a maximum of 15 out of 16 codes can be assigned for traffic. The available RRU can be obtained by:

$$N_{RRUa_{UMTS}} = (N_{co} - 1) \quad N_c - \sum_{i}^{N_{usr}} \left\lceil \frac{B_{srv,i\,[\mathrm{bps}]}}{N_{sym[\mathrm{symb/s}]} \quad M_{code\,i[\mathrm{bits/symb}]}} \right\rceil \tag{3.6}$$

where:

- N_{usr} : number of users connected to the BS;
- $B_{srv,i}$: bitrate of the i-th user connected to the BS;
- $M_{code\ i}$: modulation coding rate of the i-th user connected to the BS.

The summation parcel consists of adding the number of RRUs required from all users connected to the UMTS BS. As for GSM and LTE, one will reserve 10% of the total resources for signalling; therefore the available RRU can be defined by:

$$N_{RRUa} = N_{RRUt} \quad 0.9 - \sum_{i}^{N_{usr}} N_{RRUr,i} \tag{3.7}$$

where:

- $N_{RRUr,i}$: number of required resources by the i-th user connected to the BS;

The load of any BS can now finally be computed by (3.1), replacing the total number of RRUs, N_{RRUt}, and the number of available RRUs, N_{RRUa}, accordingly to its RAT.

The load metrics developed in this section make it possible to measure the impact of transferring load from one BS to another, via VHO, taking the differences of the RATs into consideration. Based on this framework, one can develop a load balance algorithm to be implemented in a heterogeneous radio access network.

3.1.3 Load Balance Algorithm

In order to implement a load balancing in a Heterogeneous Network (HN) environment, one can take many different approaches as shown in the state of the art section. In this thesis, load balance is studied in order to optimise network capacity and allocate the maximum number of users while respecting their QoS requirements, hence, due to the dynamic load state of a heterogeneous network, there are two main moments when it may be required to intervene: when a BS reaches its load threshold and when a new call/session request arrives. For the sake of simplicity, from now on, one will use the term “call” as any connection request from a MT to a BS, whether it refers to a voice call in CS, or a session request in PS. Furthermore, one refers to “load” as being the one defined in (3.1).

The algorithm receives as input information all the BSs status in terms of load and coverage, and the different services priority list for every RAN, previously predefined. As a result, the new incoming call or HO request, if any solution is available, is routed to one of the available RANs: GSM, UMTS or LTE.

When the LB algorithm is triggered by a new call, it runs the CAC process; if it comes from a HO request from an overloaded BS it runs the Forced Handover (FHO) routine. The former process, described in Figure 3-1, starts by sorting the reachable BSs by the MT originating the call using the Service Priority List mentioned above. Then, it checks if any BS may accept the connection without exceeding its threshold load, otherwise it runs the FHO Routine to make any BS available. The CAC algorithm ends by establishing the new connection or blocking the call/delaying it in case of unavailable resources.

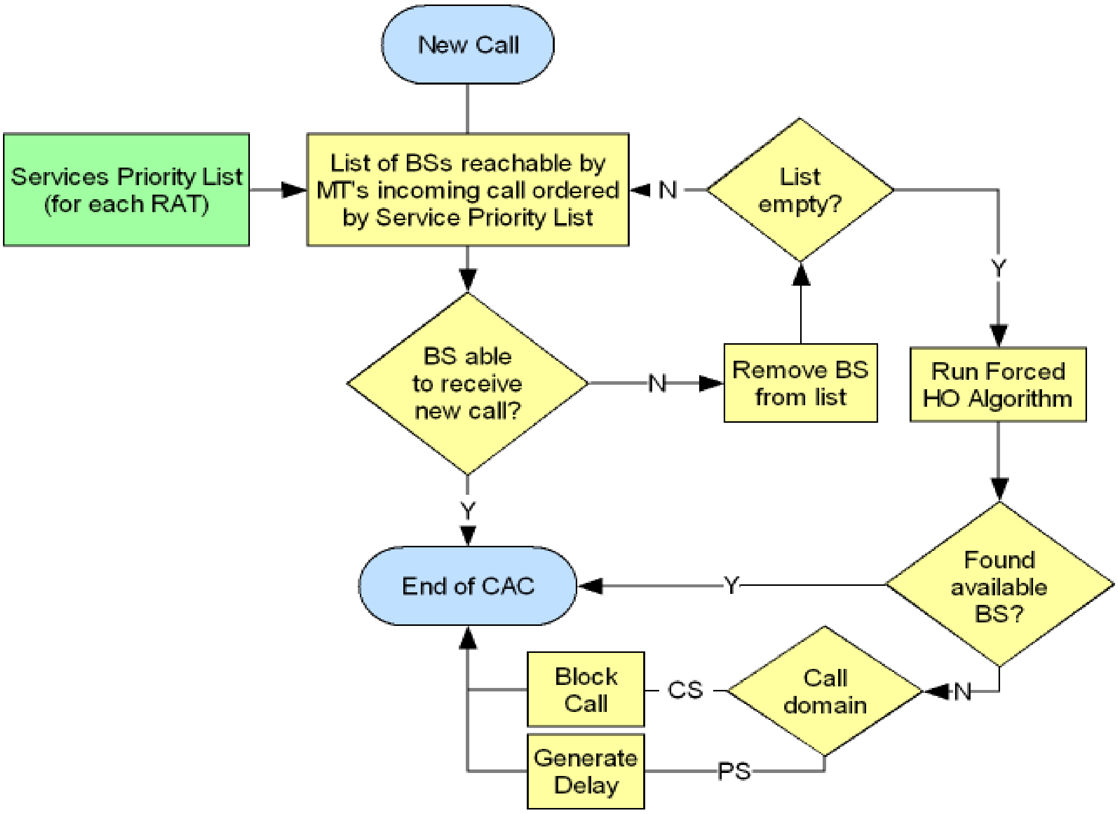


Figure 3-1 - CAC Algorithm.

A BS under the coverage of an MT is able to accept a new call only on two conditions. First, in order to receive a new call, the targeted BS has to calculate the required number of RRUs the incoming service demands (represented by index "inc") and verify if they do not surpass the number of available RRUs, as stated in:

$$N_{RRUr,inc} < N_{RRUa} \tag{3.8}$$

Thereafter, if the number of required RRUs is lower than the available ones, the new load calculated for the targeted BS to receive the new call, ρ', must not exceed a given load threshold, ρ_{thr}:

$$\rho'_{[\%]} = \left(1 - \frac{N_{RRUa} - N_{RRUr,inc}}{N_{RRUt}}\right)\ 100 < \rho_{thr\,[\%]} \qquad (3.9)$$

If both conditions are true, the BS is able to receive the call.

When the LB Algorithm is triggered by a HO request, it means that one or more BSs have reached their load thresholds. In this case, the LB Algorithm uses the FHO Routine shown Figure 3-2. The forced HO algorithm starts by ordering the BSs with the most overloaded ones on top of the list, unless the routine has been used by the CAC algorithm, where it gives priority to the BSs reachable by the MT originating the new call. Then, it follows a procedure similar to the one described in [NgAg11], where the network, by changing MTs allocation (either by HHO or VHO), tries to ease the load of the overloaded BSs. The re-allocation of MTs stops when there are no more BSs an MT can be allocated to or when there are no more overloaded BSs.

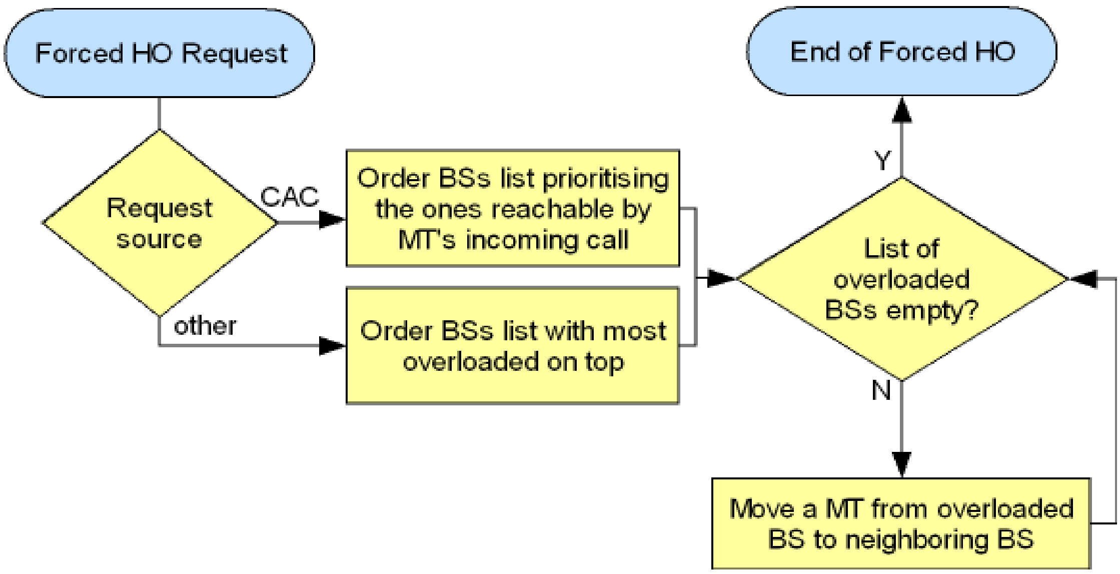


Figure 3-2 - Forced Handover Algorithm.

3.2 Simulator

3.2.1 General Description

The Common Radio Resource Management (CRRM) Simulator used to obtain the results of the present thesis is a system level, time-based simulator with a 10 ms resolution, developed by Serrador ([Serr02] and [Serr12]) over the Microsoft Visual Studio 2005 platform. The tool implements all fundamental systems functionalities, like power control, link control, basic channel code management, radio bearer service, load control, access control, propagation estimation and interference estimation and generation. Furthermore, the simulator is also capable of generating users and heterogeneous traffic services like the ones described in Section 3.1.1. Some CRRM features, such as HHO and VHO, were already implemented based on a Cost Function (CF), as well as a routine aiming at power consumption distribution among BSs.

The original version of this simulator only considered two different types of RATs, i.e., UMTS (both releases 99 and 5) and WiFi, hence, many modifications were required for the purpose of this thesis. The WiFi module was disabled during simulation time and two new dedicated modules were developed to add GSM and LTE, as represented in the CRRM Simulator block diagram in Figure 3-3. The CRRM Algorithm and Policies Engine (mainly regarding HO methods and load status computation) were also re-designed according to the algorithm and models defined in the present.

The CRRM Simulator can be divided into three main blocks, identified by the green, yellow and blue areas of Figure 3-3. The green block represents all the input information needed to be defined to run the simulator for a given scenario; the yellow block contains the RRM functionalities of the program and it is where the new models and algorithms were implemented; the blue block is where the simulator calculates and then writes the outputs into a file.

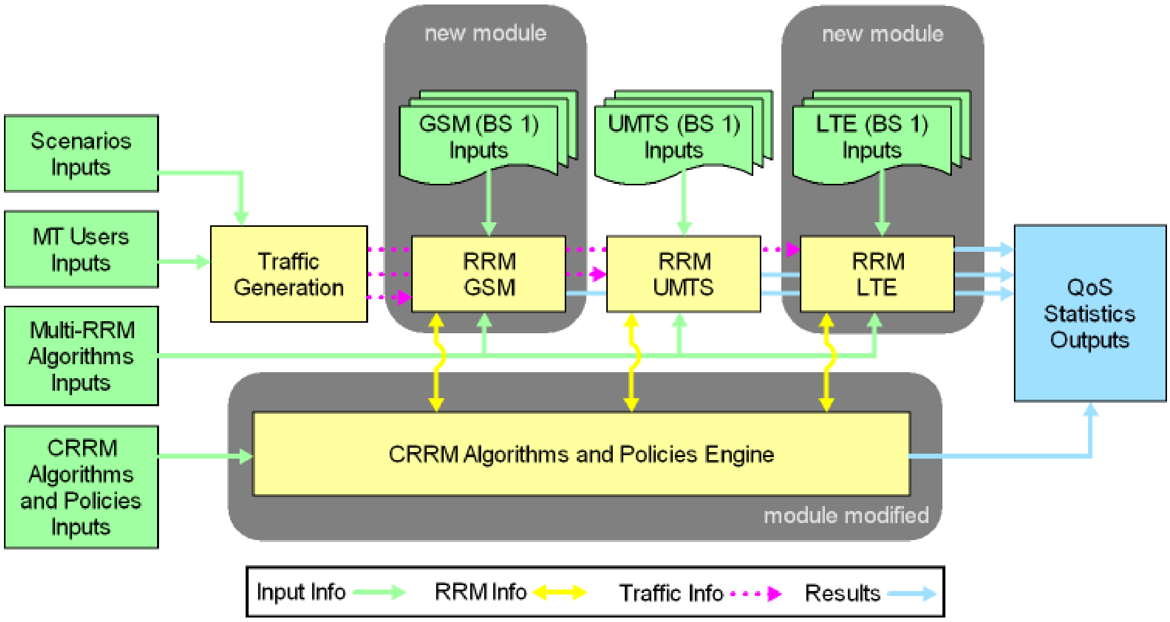


Figure 3-3 - CRRM Simulator block diagram.

The green blocks consider the following inputs:

- Scenarios Inputs: consisting of the simulation geographic area, services source models configuration data, services rates and duration, propagation models information, location of BSs, building and streets information, etc.;
- MT User Inputs: number of users, service penetration, etc.;
- Multi-RRM Algorithms Inputs: parameters related to RRM issues, like service priority list, load threshold per RAN, etc.;
- CRRM Algorithm Policies Inputs: defines the parameters related to the Cost Function weights (BS-QoS, user’s preferences), maximum QoS parameters, for each type of RAT;

- GSM, UMTS and LTE BSs Inputs: input parameters for the three different RATs for all BSs, like power level, MT maximum power, antennas pattern, total power and frequency.

The yellow set of blocks is where most of the computational effort is performed, taking a great percentage of the running time due to the recursive calculations in it:

- Traffic Generation: is the block where all traffic information vectors of all MTs and services is built, usually within a time frame of one hour;
- GSM, UMTS and LTE RRM blocks: perform the fundamental functionalities of a specific RAT, by running/managing and monitoring the radio links conditions and services attach, thus implying an intense computational effort. It is assumed that all RATs have RRM capabilities;
- CRRM Algorithms and Policies Engine: is the block where major decisions (related to the thesis theme) are taken. It is responsible for all VHOs, initial BS selection, and runs the LB algorithm for all the BSs and MTs in the scenario.

Finally, the single blue block is where output parameters are stored, most of them concerning QoS measurements and system statistics at RRM and CRRM levels.

The development of the GSM and LTE modules required a huge effort, since they required a structural change in the original simulator. Although, most of the radio features could be adapted from the existing RATs, they were spread in many parts of the simulator. The simulator consists mainly of 53 text files written in C++, 55 header files and some more auxiliary ones. All of the three blocks mentioned above had to be changed, but the main developments are highlighted in grey in Figure 3-3. The models described in Section 3.1 were translated into code at the CRRM Algorithms and Policies Engine block and the new GSM and LTE modules were written from scratch, adapting the code from the existing UMTS RAT and also by adapting the implementations done by Venes's contributions to the simulator [Vene09]. Venes has improved the simulator to add WiMAX. All RATs implement a COST231 Walfish-Ikegami propagation model and noise and interference estimation as stated in Section 3.1. Many minor routines had to be implemented or adapted from the original simulator, e.g., the one represented in Figure 3-4 to check if there are enough RRUs in a BS whenever a user is trying to connect.

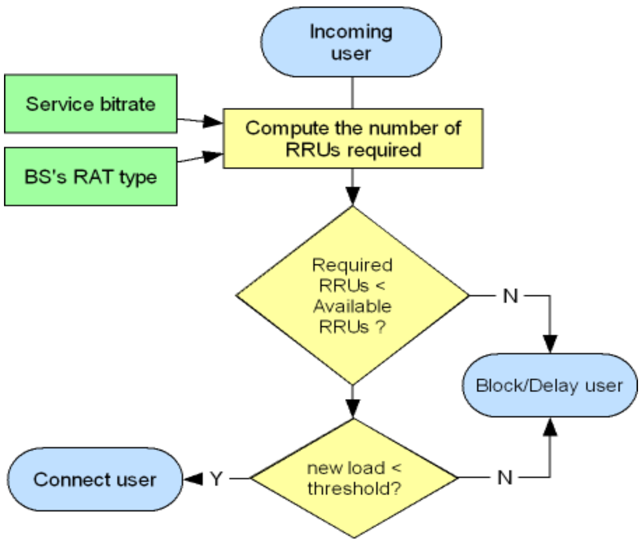


Figure 3-4 – Resource validation before accepting new calls.

The Figure 3-5 shows the Graphical User Interface (GUI) of the simulator. This user friendly environment easily allows the creation of different simulation scenarios, where one can define the number and position of BSs and MTs in an urban scenario and define a large set of their parameters.

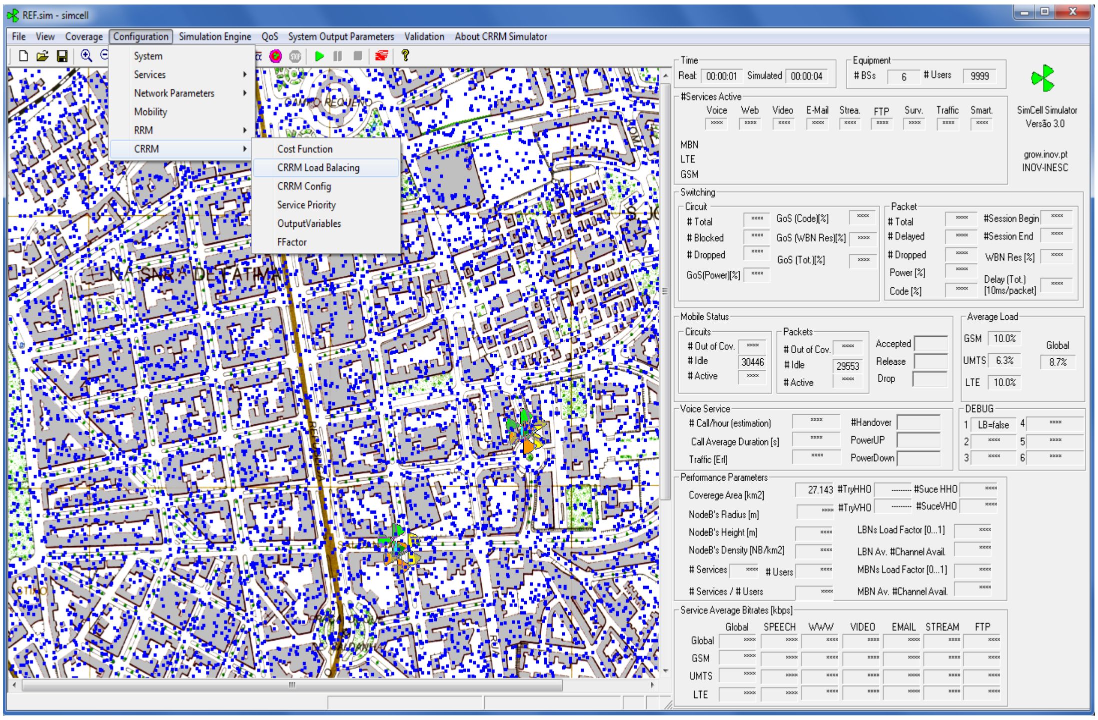


Figure 3-5 - CRRM Simulator Graphical User Interface.

3.2.2 Inputs and Outputs

The CRRM Simulator inputs, represented in green in the block diagram, are loaded from a single input file with all the required information. One can better list this information by splitting it into three groups:

- Scenarios and MT users;
- Multi-RRM Algorithms;
- CRRM Algorithms Policies

The group called “Scenarios and MT users” contains MTs specifications, services characterisation, and is where the main scenario parameters are defined, namely:

- Geographic map with buildings/streets’ disposition;
- Number of BS and of MTs;
- Numbers of fast and slow MTs;
- Minimum speeds for fast and slow MTs;
- Maximum speeds for fast and slow MTs;
- Service mix (voice, web browsing, video-telephony, e-mail, music streaming, file sharing, smart meters, eHealth, domotics and surveillance) and the corresponding individual source model parameters, described in Annex B.

The “Multi-RRM Algorithms” group defines the parameters required to define the BSs signal propagation conditions, as well as some RRM initial parameters, like:

- BS type (GSM, UMTS, or LTE), location, height, total and pilot powers, total available channels, antenna gain, orthogonal factor (CDMA), receiver noise figure, cable loss, and frequency;
- Number of channels used for signalling;
- Receiver noise density and thermal noise density;
- Building loss (penetration additional factor), height and separation distance;
- Street width and orientation;
- User body loss;
- MT maximum Tx power and height;
- Urban type;
- Services priority table (services priorities mapped into RANs).

The group named “CRRM Algorithms Policies” contains the LB algorithm rules and other load parameters, such as load threshold and load expressions defined in Section 3.1.

Regarding the CRRM Simulator output, represented in blue in the block diagram, it produces a file related to CRRM overall performance.

The simulator being structured in a modular architecture, one can define itself the statistics and output parameters that are interesting and may easily implement its extraction, by setting up virtual scopes. Most of the following output parameters were already defined and implemented, in order to reflect the network’s overall performance in a CRRM perspective.

The Blocking Probability, P_b, which is as a measure of blocked Voice and Video calls, is defined as:

$$P_b = \frac{N_{cb}}{N_c} \quad 100 \quad [\%] \tag{3.10}$$

where:

- N_{cb} : is the number of blocked Voice and Video calls by the CAC process;
- N_c : is the total number of incoming Voice and Video calls.

In the current simulator, a session can end normally while reaching its end or it may be dropped due to one of the following kinds of situation:

- HO, if the MT is not covered by a BS with enough available resources;
- In UMTS, when interference and/or load increases beyond acceptable limits;

The Average Delay, $\bar{\tau}$, is defined for Web browsing, Streaming, E-mail, FTP, Smart meters and eHealth services, as a measure of the delay affecting the transmission of packets:

$$\bar{\tau} = \frac{N_{fd}}{N_p} \quad 10 \quad [\mathrm{ms}] \tag{3.11}$$

where:

- N_{fd} : is the number of delayed frames;
- N_p : is the total number of packets transmitted.

The Drop rate, D_r, is defined in the following way:

$$D_r = \frac{N_{sd}}{N_s} \quad 100 \quad [\%] \tag{3.12}$$

where:

- N_{sd} : is the total number of dropped sessions;
- N_s: is the total number of sessions (ended normally and dropped).

The Probability of VHO failure, P_{VHOf}, is defined as the percentage of failed VHOs:

$$P_{VHOf} = \frac{N_{VHOf}}{N_{VHOa}} \quad 100 \quad [\%] \tag{3.13}$$

where:

- N_{VHOf} is the total number of failed VHO;
- N_{VHOa} is the number of VHO attempts.

The Probability of HHO failure, P_{HHOf}, is defined as the percentage of failed HHO:

$$P_{HHOf} = \frac{N_{HHOf}}{N_{HHOa}} \quad 100 \quad [\%] \tag{3.14}$$

where:

- N_{HHOf} is the total number of failed HHO;
- N_{HHOa} is the total number of HHO attempts.

Three types of Average Bitrate are defined: per service, per system and global:
The Average Bitrate per Service, B_{srv}, is defined as the average value of the bitrate for a given service.

$$B_{srv} = \sum_{i=1}^{N_{usr,srv}} \frac{B_{r,i}}{N_{usr,srv}} \tag{3.15}$$

where:

- $B_{r,i}$: bitrate of user i ;
- $N_{usr,srv}$: total number of active users in the same service type;

The Average Bitrate per RAN, B_{RAN}, is defined as the average value of the bitrate for a given RAN.

$$B_{RAN} = \sum_{i=1}^{N_{usr,RAN}} \frac{B_{r,i}}{N_{usr,RAN}} \tag{3.16}$$

where:

- $B_{r,i}$: bitrate of user i ;
- $N_{usr,RAN}$: total number of active users in the same RAN;

The Average Global Bitrate, B_G, is defined as the mean value of the Average Bitrate per RAN:

$$B_G = \frac{B_{RAN,GSM} + B_{RAN,UMTS} + B_{RAN,LTE}}{3} \tag{3.17}$$

The Average Number of BSs reachable per MT, $\overline{N}_{BSMT}$, is defined as:

$$\overline{N}_{BSMT} = \frac{1}{N_{usr}} \sum_{i=1}^{N_{usr}} N_{BSMT,i} \tag{3.18}$$

where:

- N_{usr} is the total number of users;
- $N_{BSMT,i}$ is the number of BSs reachable by user i.

The Average Load per RAN, $\overline{\rho}_{\mathrm{RAN}}$, is defined as being the average relative occupation rate of resources of all BSs for a given system, per second. This output parameter is used to plot the load evolution of a RAN system over time and is one of the main relevant indicators for the load balance performance evaluation. Its value is given by:

$$\overline{\rho}_{RAN} = \frac{1}{N_{BS}} \sum_{i=1}^{N_{BS}} \rho_i \tag{3.19}$$

where:

- $\overline{\rho}_{RAN}$: average load for a given RAN (GSM, UMTS or LTE);
- N_{BS} : number of BS of the same RAT type;

The load of any BS, ρ, is given by (3.1).

The average load per RAN is a good indicator to analyse a specific RAT performance, although one must be careful designing the simulated scenarios. For instance, considering two BSs of the same RAT may present a low $\overline{\rho}_{RAN}$, but one of the BSs may be overloaded. In order to prevent such situations, one must locate the BS in a way that they are under the same conditions in terms of covering the same amount of MTs.

3.2.3 Assessment

As the simulator used to test the previously presented models was developed by Serrador in [Serr12] and further improved by Venes in [Vene09], it has already been duly validated in the respective sources. Nonetheless, for the purpose of this thesis, new modules have been developed from scratch to add GSM and LTE BSs, as well as new traffic generation routines for M2M services that required many hours of debug simulations. Focusing on the load measurements obtained in output, one presents a brief assessment on some of the tests done to a single BS at a time. The setup in Figure 3-6 is composed of a single LTE BS operating at the 2600 GHz band with 10 MHz bandwidth covering a uniform distribution of users.

For an increasing number of users density, naturally, more users are connecting to the BS, confirmed by simulation. The results have shown that the BS's load increases linearly with the users' density as one can see in Figure 3-7. Considering that 10% of the load is reserved upfront for signalling purposes, the chart suggests that every time the number of connected users doubles, there is a 3% increase in the LTE BS's load.

In the scenario with 6000 users/km^2, the LTE BS was connected to 67 users/s on average. From theory one could expect that a single LTE BS with an average of 67 connected users/s would consume an average number of 67 RBs (considering the number of required resources per service as being 1 RRU for the bitrates presented in the reference scenario in Chapter 4). In a 10 ms period, there are a total of 1000 RBs. Taking 10% out for signalling purposes, one gets 900 RBs available for communications which, after subtracting the referred 67 RBs, leads to an expected load of 16.7 % according to (3.1). By simulation, one has obtained an average load of 15.6 % for the same parameters, which is a pretty close value, considering all the random parameters involved in the simulator.

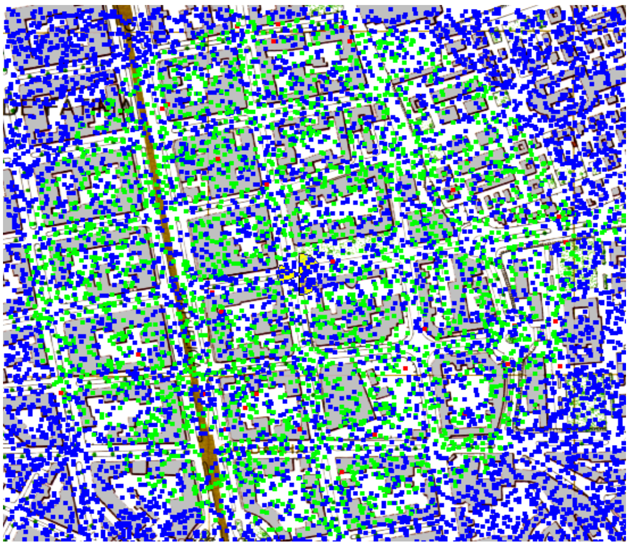
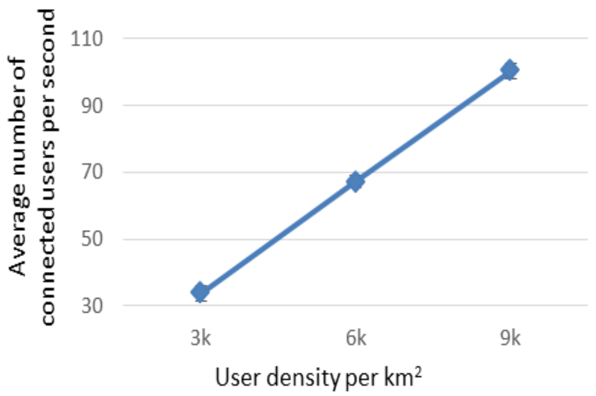
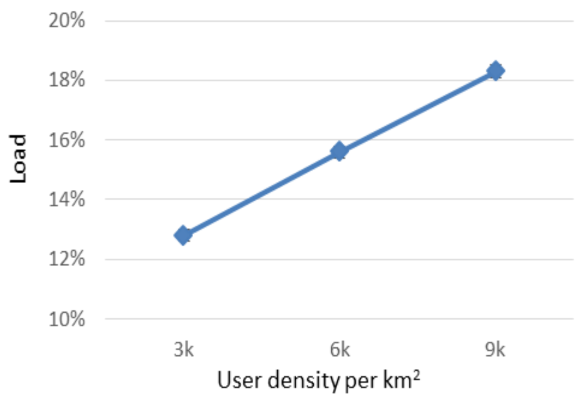


Figure 3-6 - Assessment scenario for a single LTE BS.

A similar procedure was made to both GSM and UMTS BSs whose results obtained in simulation can be seen in Annex D.



(a) Number of connected users



(b) Load

Figure 3-7 - LTE Assessment.

To infer on the quality of the simulator outputs, the expected values and the simulation results are summarised in Table 3-1, considering for each case the same average number of connected users. The variation between expected values and simulated ones is always below 10%.

Table 3-1 - Validation of the simulation's output against the expected values.

	Expected load (theory) [%]			Load (via simulation) [%]			Difference [%]			Variation, Δ [%]		
	GSM	UMTS	LTE	GSM	UMTS	LTE	GSM	UMTS	LTE	GSM	UMTS	LTE
3k	62.5	20.1	13.4	63.0	18.6	12.8	0.5	1.5	0.6	0.8	7.5	4.5
6k	70.0	33.2	16.7	69.8	30.4	15.6	0.2	2.8	1.1	0.3	8.4	6.6
9k	71.3	47.1	20.0	71.2	42.8	18.3	0.1	4.3	1.7	0.1	9.1	8.5

The SimCell simulator inherited the 10 minutes initial setup variation from previous versions, therefore, its values are discarded. Furthermore, due to resource allocation modifications, some changes to the traffic generation were required, which lead to an abrupt traffic loss in the last 10 minutes, as plotted in Figure 3-8. In order to ensure all measurements are within a constant traffic profile, the last 15 minutes of simulation are also discarded leading to a total useful simulation time of 45 minutes of real time traffic. Load variations in Table 3-1 should become even closer to zero once the first 10 and last 15 minutes of simulation are discarded, since the values in this time window are not trust worthy.

In terms of output data, it was only required to add new load information to the original output block, hence, one performed some test to infer on its variation and convergence values for good measurements. The setup scenario has been identically to the one used to assess LTE, except this time with 3 different RATs collocated, and discarding the last 15 minutes of simulation; Figure 3-9 shows the load evolution of a single simulation.

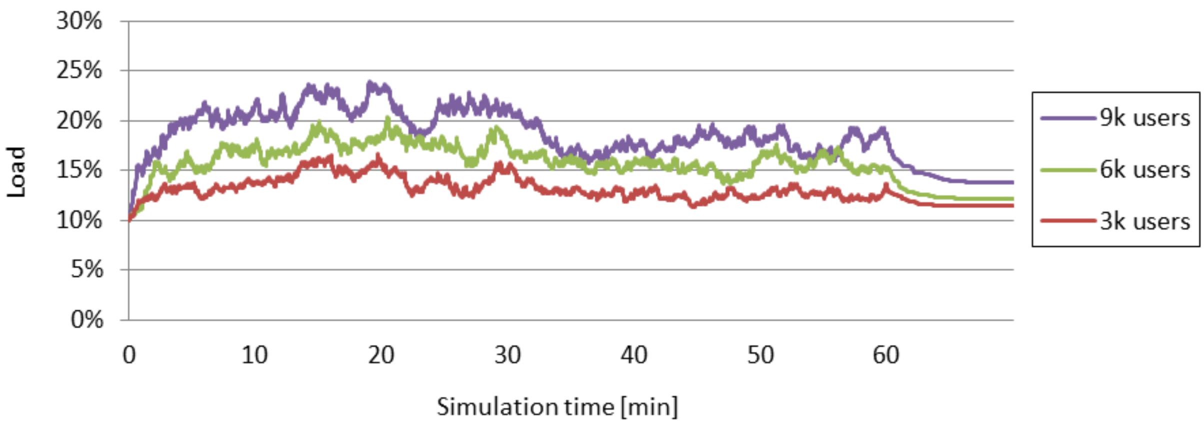


Figure 3-8 - LTE Assessment - Load time evolution.

The same scenario was simulate 15 times to analyse the HNP (Heterogeneous Network Performance) of the load in the 3 different RATs defining a measure of convergence, Δ, as follows:

$$\Delta[\%] = \frac{|HNP^{cum,s} - HNP^{cum,S}|}{HNP^{cum,S}} \quad 100 \tag{3.20}$$

where:

- s : simulation run index;
- S : total number of simulation runs conducted during the convergence study (equal to 15);
- $HNP^{cum,s}$ is the partial HNP cumulative mean at simulation run s;
- HNP^{cum_S} is the total HNP cumulative mean (from all simulations).

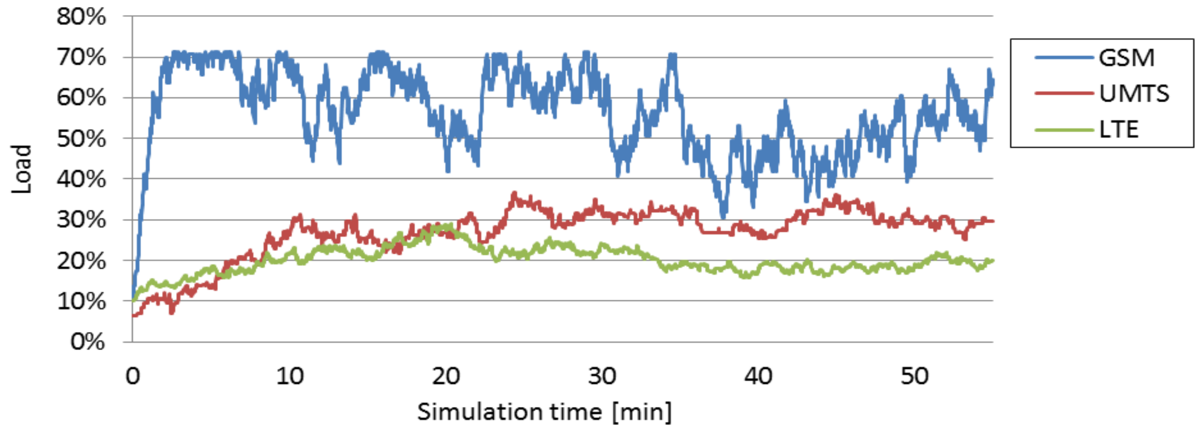


Figure 3-9 - Load time evolution for the 3 RATs.

As one can observe in Figure 3-10, its load variation converges to less than 10% variation in just 4 simulation runs for all 3 RANs.

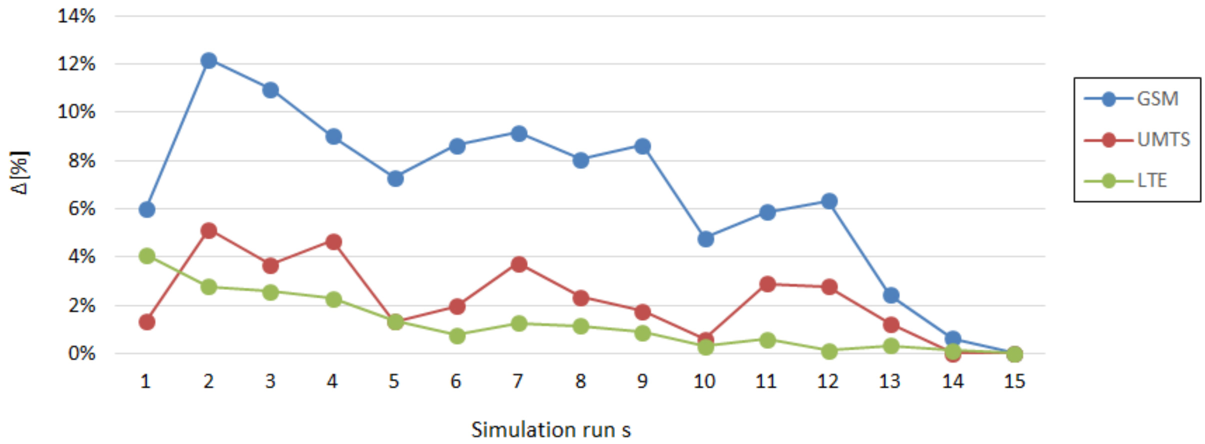


Figure 3-10 - Convergence analysis.

Previous versions of the simulator also required 5 simulations per scenario being tested to obtain a variation below 10%, so it has been decided to keep the same number of simulation runs to get the mean value and its standard deviation.

The assessment concludes that all new modules added to the simulator are working as expected. Five simulations were ran for every scenario with a duration of 55 minutes, whose initial 10 minutes were discarded. 12 scenarios were tested, leading to a total number of 60 simulations. Each simulation run takes on average 2 hours of computational effort, implying a total of 120 hours, which is an acceptable value within the time constraints of a master thesis.

Chapter 4

Analysis of Results

This chapter contains the scenarios tested and results obtained via simulation for the algorithms and models previously presented. In the first section, the reference scenario is described, followed by the analysis of the results in the second section.

4.1 Reference Scenario

The Reference Scenario (REF) to evaluate the load balance performance of the models previously described in Chapter 3 is an urban environment supporting multiple RATs, i.e., OFDMA (LTE), CDMA (UMTS), and FDMA/TDMA (GSM). It was designed in such a way that it allows the simulation of a variety of different (sometimes unrelated) parameters and yet being possible to compare the load impact among those scenarios. Therefore, REF considers only 3 co-located BSs, 1 per RAN, Figure 4-1. This approach avoids VHOs that would mask the real impact on the load of the BSs and also focus the traffic within the coverage area around the site.

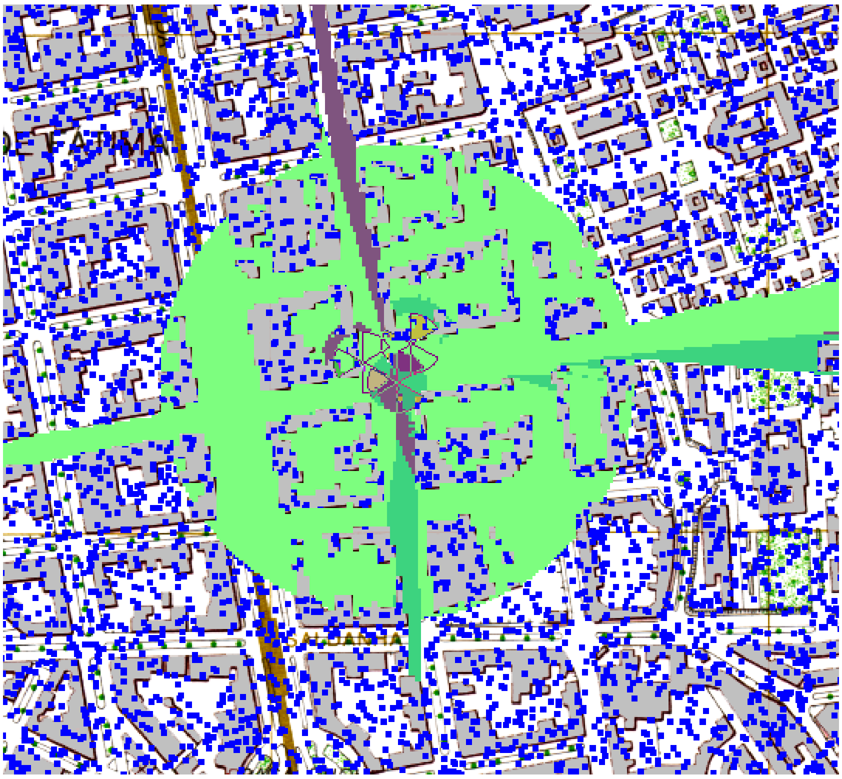


Figure 4-1 - Reference scenario.

The main characteristic of the BSs are the following:

- The OFDMA RAT is based on 10 MHz bandwidth LTE operating at the 2.6 GHz band. Each cell has 100 RRUs, which can be assigned to traffic bearers.
- The configurations for CDMA cells are chosen according to UMTS (HSDPA) working at 2.1 GHz. Each cell has 3 carriers, and each carrier has 16 codes. Only 45 codes out of all 48 codes can be assigned to users' traffic.
- The biggest cell size is configured for FDMA/TDMA, based on GSM900, each cell having 10 radio channels and each channel having 8 timeslots. It is assumed that 75 timeslots out of a total of 80 available ones can be used for users' traffic.

The time resolution of the simulator is 10 ms, which is not compatible with the TTI of the RANs, so, a normalisation had to be done. In GSM an operator may use 10 frequency channels, each one carrying 8 TS every TTI of 4.61 ms. An approximation to a TTI of 5 ms allowed to double the amount of RRUs

GSM produces in any simulation step so, instead of 80 RRUs, GSM has a total of 160 RRUs per 10 ms. The assumption of 48 codes available per cell in UMTS (HSDPA) is valid for a TTI of 2 ms and to comply with the simulator's time resolution, this number of available RRUs is multiplied by 5 leading to a total number of 240 RRUs. A similar normalisation has been performed in LTE although, in this case the TTI is 1 ms, and considering that in any 0.5 ms RBs are generated, it is necessary to multiply the RRUs by 20 to get the number of RRUs available any 10 ms. Moreover, the number of RB is bandwidth-dependent, and one must consider that the 10MHz bandwidth leads to a final number of 1000 RRUs per 10 ms (50 RB multiplied by 20). The relevant parameters of the 3 BSs are summed up on Table 4-1.

Table 4-1 – Resource normalisation for the simulator time resolution.

	GSM	UMTS	LTE
Frequency band [MHz]	900	2 100	2 600
Channel bandwidth [MHz]	0.2	5	10
TTI [ms]	4.61	2	1
Number of RRUs in a TTI	80	48	100
Number of RRUs in 10 ms	160	240	1 000

Regarding the propagation models, the values for the COST 231 Walfish-Ikegami that better characterise REF are shown in Table 4-2, being kept the same for all simulated scenarios.

Table 4-2 - COST 231 Walfish-Ikegami Model parameters.

Height of the BS antenna, h_b [m]	22
Height of the buildings, H_B [m]	25
MT height, h_m [m]	1.5
Street width, w_s [m]	20
Building separation, w_B [m]	40
Incidence angle, ϕ [°]	90

Users are characterised by their position and service. Regarding their position, in REF, users are considered to be static and randomly located according to a bi-dimensional uniform distribution in a (1000 × 1000) m^2 area with a density of 6 000 km^2, which corresponds exactly to 6 000 users. In order to simplify REF, and ease the required computational effort, only one type of service is allowed per user during the whole simulation period. There are 10 kinds of services allocated to users at simulation start, implemented according to the source models described in Annex B, and respecting the penetration ratios of Table 4-3 and a service priority scheme defined on Table 4-4.

Table 4-3 - Service characterisation for REF.

	Service penetration [%]	Size [kB]	Duration [s]	Service bitrate [kbps]		
				Min	avg	Max
Voice	47.0	-	60	12	32	64
WWW	10.0	180	-	100	1 024	-
Video	3.6	-	-	384	1 024	2 048
E-mail	9.0	300	-	100	600	-
Streaming	18.0	-	90	128	2 048	3 000
FTP	8.1	2 000	-	384	2 048	4 096
Smart Met.	1.0	2.5	-	-	200	-
eHealth	0.9	5 480	-	-	200	-
Domotics	0.4	2.5	-	-	200	-
Surveillance	2.0	5.5	-	128	200	2 048

Table 4-4 - REF - Service priority.

RAN / **Service**		GSM	UMTS	LTE
Voice Call		1	2	3
Web Browsing		3	2	1
Video Call		3	2	1
E-Mail		3	1	2
Music Streaming		3	2	1
File Sharing		3	2	1
M2M	Smart Meters	3	1	2
	e-Health	3	2	1
	Domotics	3	2	1
	Surveillance	3	2	1

Any time a user starts a new service, a bitrate is randomly established, its value remaining fixed during a single simulation step of 10 ms, whose minimum, maximum and average bitrates values are also defined for REF in Table 4-3. Once the service bitrate is set, the CRRM module tries to connect the user to the network by finding a BS under its coverage that guarantees both the required QoS and the overall network efficiency. For that matter, the CRRM implements a service priority list for each RAT, defined for REF as stated in Table 4-4, designed on the following assumptions:

- voice calls may be steered to GSM with maximum (1) priority;
- background traffic, like e-mail and smart meters, is served by UMTS due to lower delay demands in terms of QoS comparing to the rest of the services;
- all remaining services must prioritise their connection to an LTE BS.

This framework intends to characterise a realistic approach of a normal day traffic in an urban scenario, whose simulations represent 45 minutes of real life traffic. The most relevant results coming from the simulator's output are in Table 4-5, being used to compare load performances in the following sections.

However, it is important to note that, despite all parameters having been defined for the reference scenario, some parameters may present, on simulation, slight variations from the defined ones, due to the random variables associated with a radio environment, such as interference, user's location and traffic generation. For instance, the average service penetration obtained through simulation shown in Figure 4-2 is slightly different from the one defined in Table 4-3.

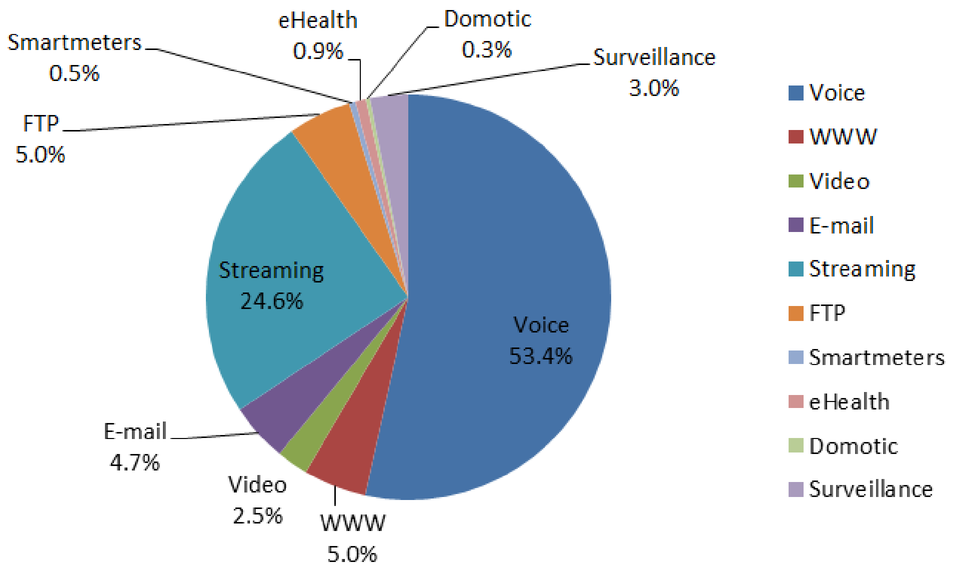


Figure 4-2 - REF - Service penetration obtained in simulation.

Therefore, all the results obtained through simulation must be duly analysed, not only by respecting the statistically relevance imposed by the assessment conclusions, but also in terms of cross checking the expected values with the ones defined for simulation, and the ones really obtained from simulation.

The results in Table 4-5 also suggest a higher load value for GSM compared to the one that would be predicted, due to voice calls being steered to GSM with a 53.4 % service penetration ratio. Furthermore, this is the RAT that has a lower number of RRUs, with lower capacity per RRUs.

Table 4-5 – REF - Network performance outputs.

Simulation Output		Mean val.	St. dev.
Blocking Probability, P_b [%]		0.40	0.48
Delay, τ [ms]		0.78	1.03
Drop rate, D_r [%]		0.0	0.0
Average Connected Users per second	GSM	33.7	2.5
	UMTS	5.4	0.7
	LTE	12.6	1.5
Average Load, $\bar{\rho}$ [%]	GSM	52.4	3.0
	UMTS	10.9	1.3
	LTE	13.7	0.6
Average number of connected users per second	Voice	29.0	2.5
	WWW	2.7	0.6
	Video	1.4	0.3
	E-mail	2.5	0.3
	Stream	13.4	1.3
	FTP	2.7	0.6
	Smart meters	0.3	0.1
	eHealth	0.5	0.6
	Domotic	0.2	0.1
	Surveillance	1.7	0.5

4.2 Load Performance Analysis

In order to test the performance of the load balance algorithm in a heterogeneous network, many scenarios have been simulated. The present section studies the impact of user and network parameters to draw conclusions on the load balance effectiveness in improving systems' overall performance. The tests described below are analysed against the reference scenario of Section 4.1.

4.2.1 Impact of Users' Density

The impact of the number of users on the load was tested for REF just varying the users' density per km^2. The simulation plotted in Figure 4-3 and Figure 4-4 shows a slight increase of load for both UMTS and LTE. The reason for such low load increases is due to more than half of the traffic being for voice calls, that is being steered to GSM BSs, and because, for the remaining packet services, LTE does not suffer a great impact on resource allocation with a 10 MHz bandwidth where RRUs for LTE are abundant for the current traffic demands.

For the GSM BS, on the other hand, one can observe a high correlation between users' density and load state. For instance, the increased variation from 3 000 users/km^2 to 6 000 users/km^2 represents a 25% increase in the load for GSM. It is also notably that for REF, an average of 34 users connected to the GSM BS are consuming 52% of the BS's total resources.

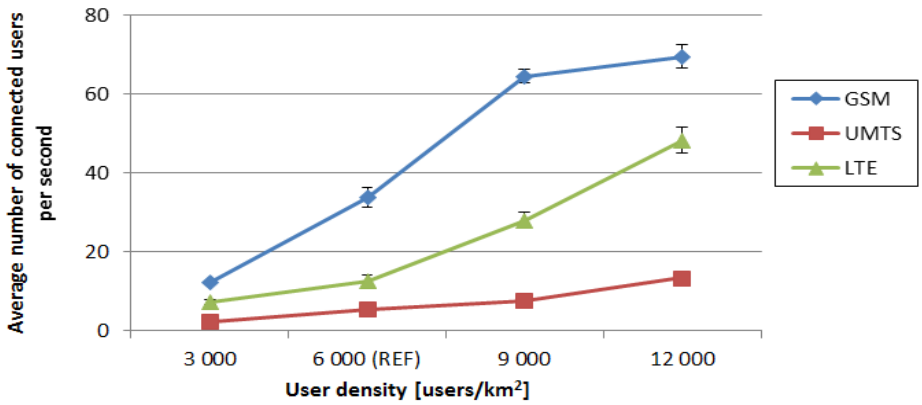


Figure 4-3 - Impact of users' density on the average number of connected users.

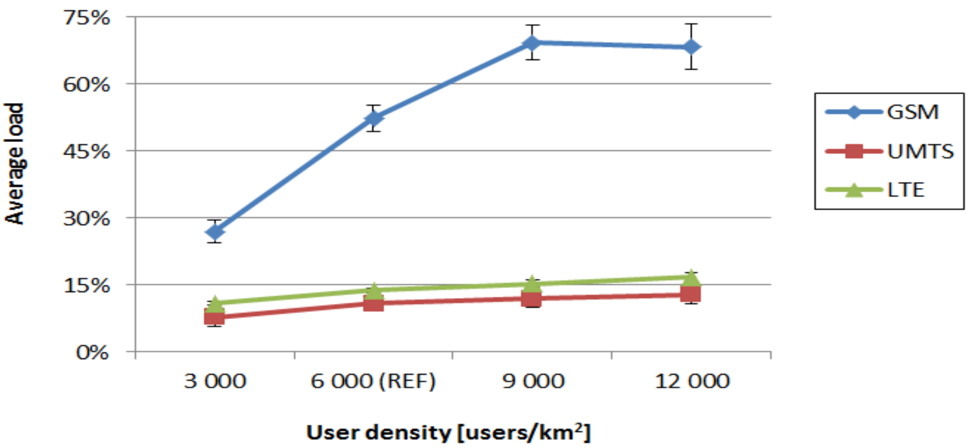


Figure 4-4 - Impact of users' density on average load.

At 9 000 users/km^2, GSM reaches the load threshold, defined for 70% and starts blocking voice calls, as shown in Figure 4-5. The blocking probability parameter counts all the calls for voice and video that are blocked within the coverage of the REF site.

The measurements for delay and drop are not relevant for the analysis of this scenario, since it only affects the performance of GSM, mainly performing voice calls. The most relevant results obtained in simulation are summed up in Table 4-6 and Figure 4-6.

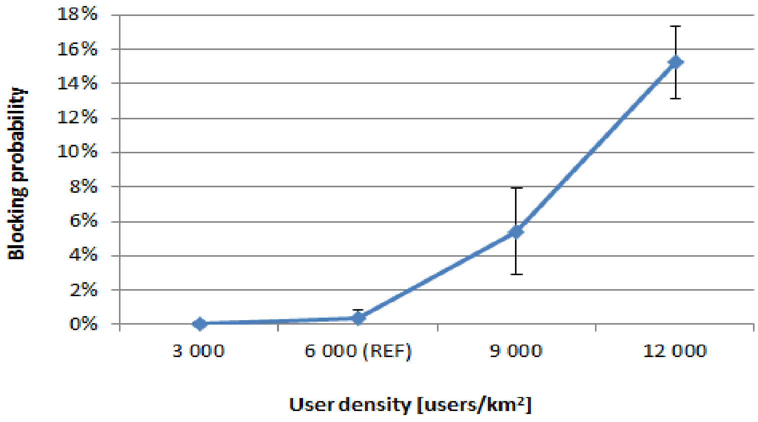


Figure 4-5 - Impact of users' density on blocking probability.

Table 4-6 – Impact of users' density on load.

Simulation Output		User density [users/km^2]							
		3k		6k (REF)		9k		12k	
		Mean	**Dev.**	**Mean**	**Dev.**	**Mean**	**Dev.**	**Mean**	**Dev.**
Blocking Probability, P_b [%]		0.1	0.0	0.4	0.5	5.4	2.5	15.2	2.1
Average Connected Users per second	GSM	12.2	0.4	33.7	2.5	64.5	1.8	69.5	2.9
	UMTS	2.3	1.0	5.4	0.7	7.6	1.1	13.3	1.2
	LTE	7.2	0.5	12.6	1.5	27.9	2.1	48.3	3.2
Average Load, $\bar{\rho}$ [%]	GSM	26.9	2.5	52.4	3.0	69.2	4.0	68.3	5.0
	UMTS	7.7	2.0	10.9	1.3	12.0	2.1	12.9	2.0
	LTE	10.9	0.5	13.7	0.6	15.3	0.9	16.8	0.9

The main conclusions drawn from the analysis of the variation of the number of users is that it hardly affects the availability of UMTS and LTE, due to the amount of RRUs these two systems have for traffic. Thereafter, it is necessary also to study the impact of changing the radio configuration parameters of LTE to be possible to analyse the impact on load. The following section studies this in further detail.

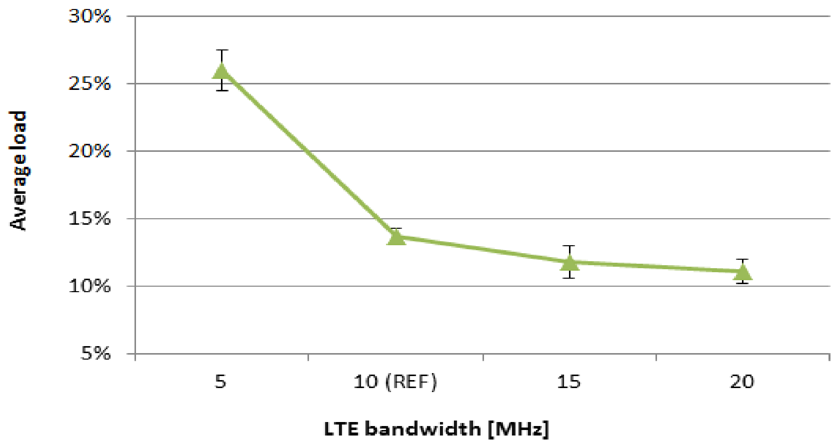


Figure 4-6 - Impact of LTE bandwidth variation on the BS's load.

4.2.2 Impact of Radio Parameters

In order to test the impact of LTE radio parameters, one has simulated the same heterogeneous reference scenario with a single GSM, UMTS and LTE BSs, varying just LTE's bandwidth for 5 MHz, 10 MHz, 15 MHz and 20 MHz. The results show little or no impact on the average number of connected users in all 3 RATs. Once the user density is defined, the number of connected users per system is always the same, apart from a small (yet negligible, over 5 runs of simulations) variation due to randomness of the simulator traffic/users generation.

The bandwidth variation of the LTE BS only impacts on LTE BS's load as plotted in Figure 4-6 from the results in Table 4-7.

Table 4-7 - Impact of LTE's bandwidth variation on load.

Simulation Output	Bandwidth [MHz]							
	5		10		15		20	
	Mean	Dev.	Mean	Dev.	Mean	Dev.	Mean	Dev.
Average Connected Users per second	11.8	1.0	12.6	1.5	14.5	1.0	12.6	1.5
Average Load, ρ [%]	26.0	1.5	13.7	0.6	11.8	1.2	11.1	0.9

Based on the information provided in Chapter 3, one can preview the number of RRUs available in LTE for each bandwidth scenario and consequently understand the great impact it creates by changing LTE's bandwidth. Bandwidth allocation is directly related to the number of RBs generated in LTE leading to the total RRUs in Table 4-8, recalling that the TTI in LTE is 1 ms and that the simulator time resolution is 10 times greater.

Table 4-8 - RRUs variation with bandwidth.

	Bandwidth [MHz]					
	1.4	3	5	10	15	20
RB [per 0.5ms]	6	15	25	50	75	100
Number of RRUs in a TTI	12	30	50	100	150	200
Number of RRUs in 10 ms	120	300	500	1 000	1 500	2 000

Considering (3.5), knowing that in a normal CP there are 7 OFDM symbols per RB, and assuming a constant modulation coding rate of 4 bits per symbol, the number of required resources any service demands is given by:

$$N_{RRUr_{LTE}} = \left\lceil \frac{B_{srv\ [\text{bits per 10ms}]}}{N_{RB_{[\text{per 10ms}]}} \quad 28_{[\text{bits}]}} \right\rceil \tag{4.1}$$

One can analyse the average consumption of RRUs by average services' bitrates of REF defined in Table 4-3 for the number of RRUs of every bandwidth. The results coming from (4.1) are presented in Table 4-9. Before applying the equation, it is necessary to consider that the total number of RB, N_{RB}, should be multiplied by 0.90 to reserve 10% of resources for signalling purposes

Table 4-9 - Number of required RRUs in LTE for FTP and M2M services.

		Bandwidth [MHz]					
		1.4	3	5	10	15	20
Number of RRUs in 10 ms		120	300	500	1 000	1 500	2 000
Number of required RRUs	Voice	1	1	1	1	1	1
	WWW	4	2	1	1	1	1
	Video	4	2	1	1	1	1
	E-mail	2	1	1	1	1	1
	Streaming	7	3	2	1	1	1
	FTP	7	3	2	1	1	1
	Smart meters	1	1	1	1	1	1
	eHealth	1	1	1	1	1	1
	Domotic	1	1	1	1	1	1
	Surveillance	1	1	1	1	1	1

The number of required resources by each service has a larger impact on the load state of the LTE BS the lower the bandwidth is. Figure 4-7 illustrates the impact of changing the bandwidth in LTE. For instance, for the 1.4 MHz bandwidth scenario, it is only possible to serve 15 users simultaneously in a file sharing service.

Other radio parameters, such as the number of radio channels allocated for communications in GSM or the number of codes given in UMTS, were kept unchanged during simulation, but the results also show how such slight changes in the radio parameters cause drastic changes in the load state of BSs.

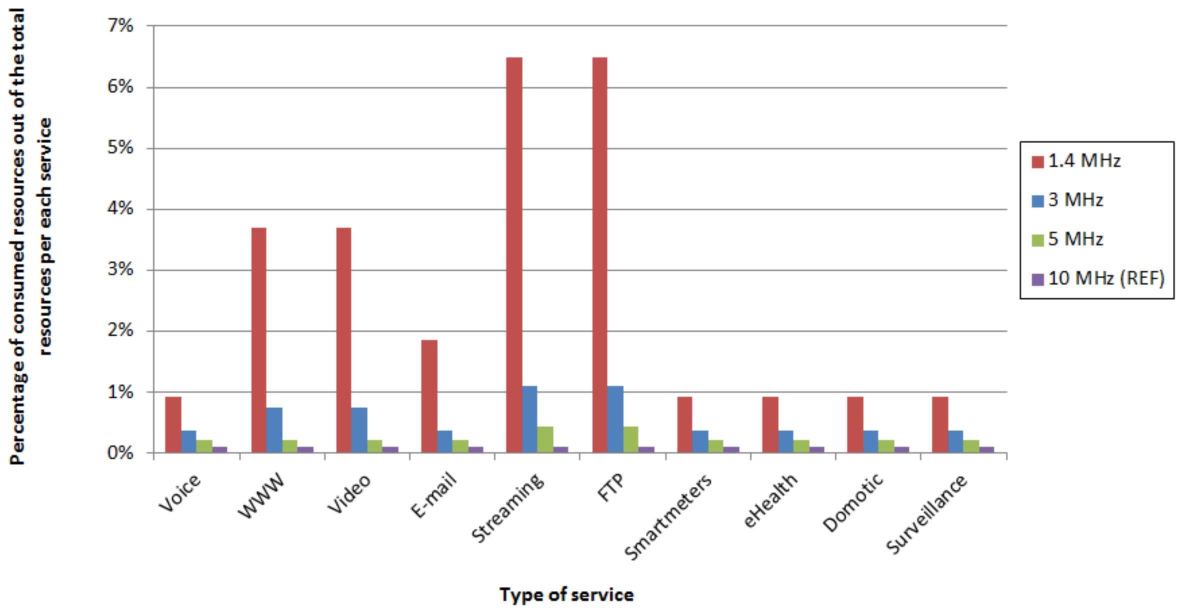


Figure 4-7 - Impact of services in the total radio resources consumption in LTE.

4.2.3 Variation of Number of Base Stations

Another scenario designed to test load variation considered the addition of more BSs in REF (“1 site”). For this case, two situations were studied: the impact of adding another site with all 3 RATs collocated (named “2 sites”) and the scenario where only one LTE BS is added to REF (“1 site + LTE”). The new scenarios are depicted in Figure 4-8 and Figure 4-9, being designed in such a way that BSs are all within the same coverage area to infer on the impact of load distribution among the different RATs.

The results, shown in Table 4-10, suggest a general increase in the number of connected users and a general decrease in the load state of the BSs. This would be expected, since the addition of a BS provides more coverage area, increasing the number of users, and more availability of resources for traffic cutting the average load of a given RAT. Nonetheless, some results differ from this description. When one single LTE BS is added, for instance, there is a decrease in the number of connected users in GSM and UMTS. This result is rather interesting in terms of easing other systems and virtual sharing the load even without applying any load balancing model.

With an increased LTE coverage, more users try to connect to this RAT (the REF’s priority table for packet services steer most of the traffic to LTE), thereby leading to less users connecting to GSM, and consequently leaving more free resources in UMTS to accept voice and video calls, easing in the

GSM load state.

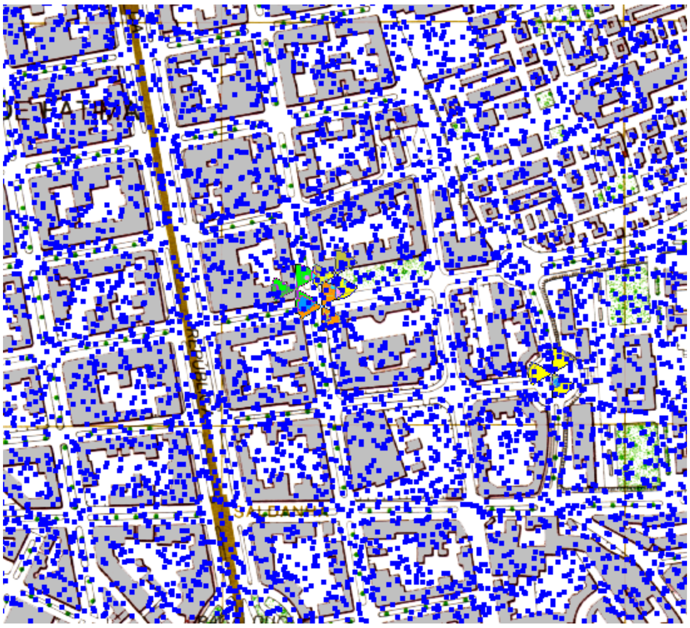


Figure 4-8 - Scenario “1 site + LTE”.

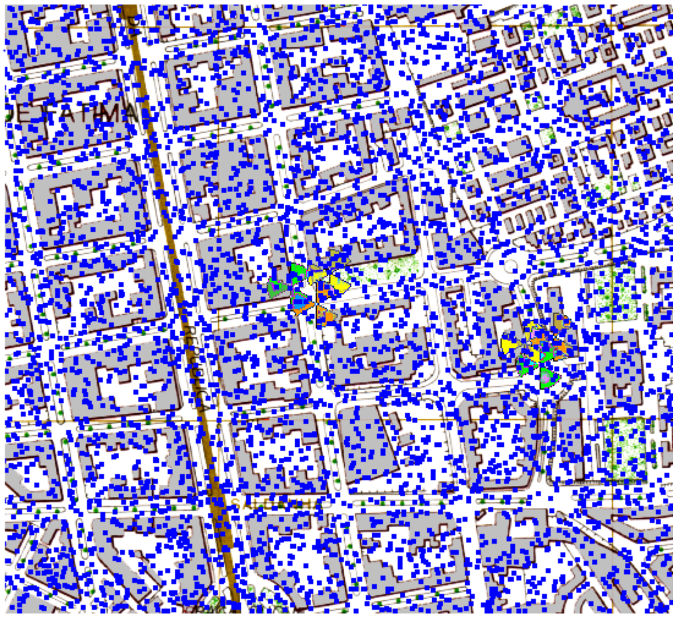


Figure 4-9 – Scenario “2 sites”.

The results obtained from simulation are shown in Figure 4-10 and Figure 4-11. This technique of adding more BSs is a common way to provide more capacity and to reduce the load in real live networks.

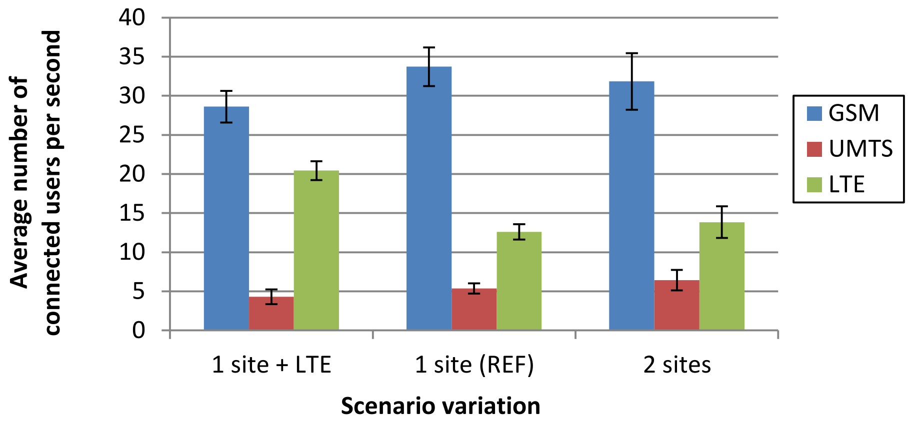


Figure 4-10 – Impact of the number of RATs on the number of connected users.

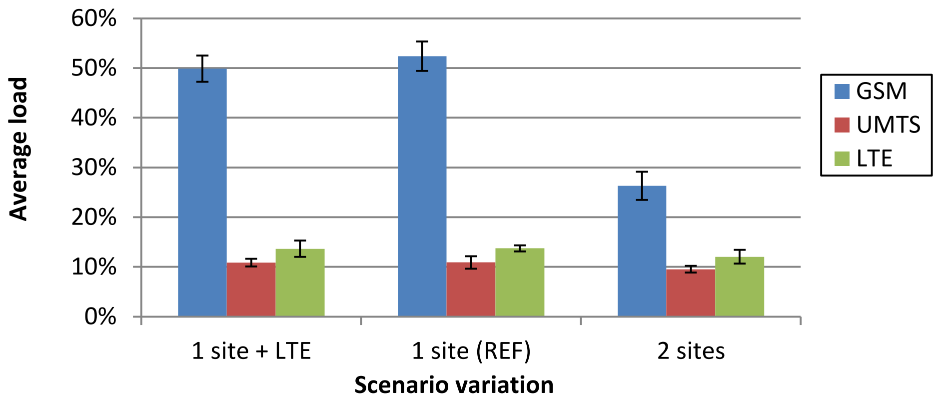


Figure 4-11 – Impact of the number of RATs on the BS's load.

Table 4-10 - Performance parameters for variation on the number of RATs in the scenario.

		Scenario					
		1 site + LTE		**1 site**		**2 site**	
Simulation Output		**Mean**	**Dev.**	**Mean**	**Dev.**	**Mean**	**Dev.**
Average Connected Users per second	GSM	28.6	2.0	33.7	2.5	31.8	3.6
	UMTS	4.3	0.9	5.4	0.7	6.4	1.3
	LTE	20.4	1.2	12.6	1.0	26.3	2.8
Average Load, ρ [%]	GSM	49.9	2.6	52.4	3.0	26.3	2.8
	UMTS	10.8	0.8	10.9	1.3	9.5	0.7
	LTE	13.6	1.6	13.7	0.6	12.0	1.4

4.3 Load Balance Model Performance

The analysis from previous sections suggests that a slight variation of radio resources has a large impact on the average system load. One can also infer that a great number of users will easily saturate the network's capacity, hence, in order to evaluate the load balance developed models, the scenario summarised in Table 4-11 was first simulated without load balance routine active in order to draw conclusions. The load balancing testing scenario consists of the collocation of the 3RATs in two sites within an urban scenario of 10 000 users/km^2. The scenarios analysed so far suggest that the BSs hardly reach half of its load capacity with the current setup, so the radio resources had to be reduced. For this purpose, the number of carriers in UMTS has been reduced from 3 to 2, cutting the available RRUs for traffic from 225 to 145; in LTE, instead of considering a 10 MHz bandwidth allocation, now one considers only 5 MHz, reducing the number of RRUs available for traffic to only 450, instead of 900. The results from previous sections also highlight the large consumption of resources in GSM due to the service penetration of voice calls, so, in order to reduce the user's allocation in GSM to better analyse the model, the service penetration was also redefined for the values in Table 4-11. All other parameters not mentioned here remain as defined before for REF.

Table 4-11 - Parameters changed from REF to test the load balance model.

Simulator Parameter		**Value**
User density [users/km^2]		10 000
Number of carriers in UMTS		2
LTE's bandwidth [MHz]		5
Service penetration [%]	Voice	35.0
	WWW	15.0
	Video	3.6
	E-mail	15.3
	Streaming	18.0
	FTP	8.1
	Smart Met.	1.7
	eHealth	0.9
	Domotics	0.4
	Surveillance	2.0

The obtained results, although increasing the expected load state per BS, do not reach the defined load threshold of 70%, therefore showing little or no improvement in the overall performance of the load balance model developed. Nonetheless, the blocking probability performance parameter is reduced from 0.6% to null, as plotted in Figure 4-12. This result shows some improvement with the LB model activated, since the CAC algorithm designed is now capable of finding more BS within reach (regardless of its RAT) increasing the odds of successful connections before blocking an incoming call.

Nonetheless, the blocking probability improvement is not that interesting, considering that the scenario without LB active registers a blocking probability below the reference QoS value of 1%, so in order to test the efficiency of the model developed, one must reach the load threshold to infer the behaviour of the BSs in the situation of resources scarcity. Such scenario could be designed by increasing the user density around the site, recreating, for example, a music festival event in a real life network, but such scenario would demand lots of time and a huge computational effort. In order to avoid such scenario, the high load demand has been virtually induced in all BSs by changing their load threshold to 25%.

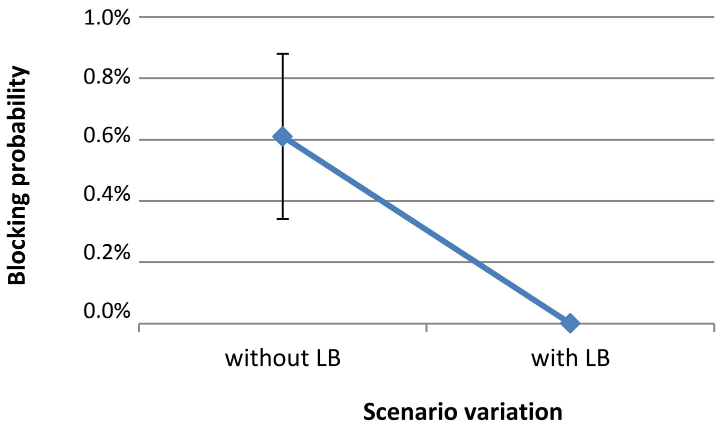


Figure 4-12 - LB model performance for blocking probability.

The results are shown in Figure 4-13 to Figure 4-17, whose scenarios description for LB and “NoLB” refer to the scenario described above and the scenario “Threshold 25% LB” and “Threshold 25% NoLB” refer to the same scenario just considering a 25% threshold value.

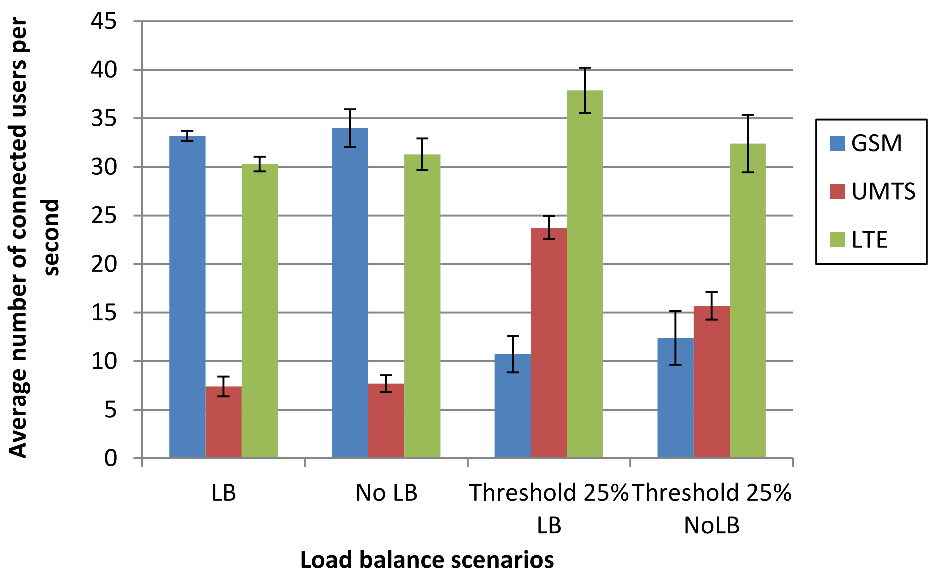


Figure 4-13 - LB model impact on the average number of users connected.

The results for a threshold of 25% show that, without LB, some voice users, after the GSM BSs have reached their load threshold, try to connect to UMTS, increasing both the average number of users connected to UMTS and increasing the UMTS BSs load state, until they also reach the load threshold.

LTE BSs are not affected in this scenario by the load threshold in case there is no LB algorithm running. On the other hand, one observes that, with LB, LTE increases its number of connected users, almost reaching the threshold. As for UMTS, the activation of the LB routine saturates all of its RRUs changing the load state from 20% to 25%, increasing the average number of connected users by 8. Moreover, without LB, GSM blocks most of the calls in the admission control, which is why the blocking probability is 15.2%, against the LB scenario where this parameter is 3.4%. Regarding the delay parameter, all 4 scenarios register a value much lower than 30ms, which is much lower than the expected one, since the overload conditions predict higher delays in the transfer of packets. Regarding the drop rate, LB presents an effective improvement in the overall RAN reducing it from 2.3% to 1.2%.

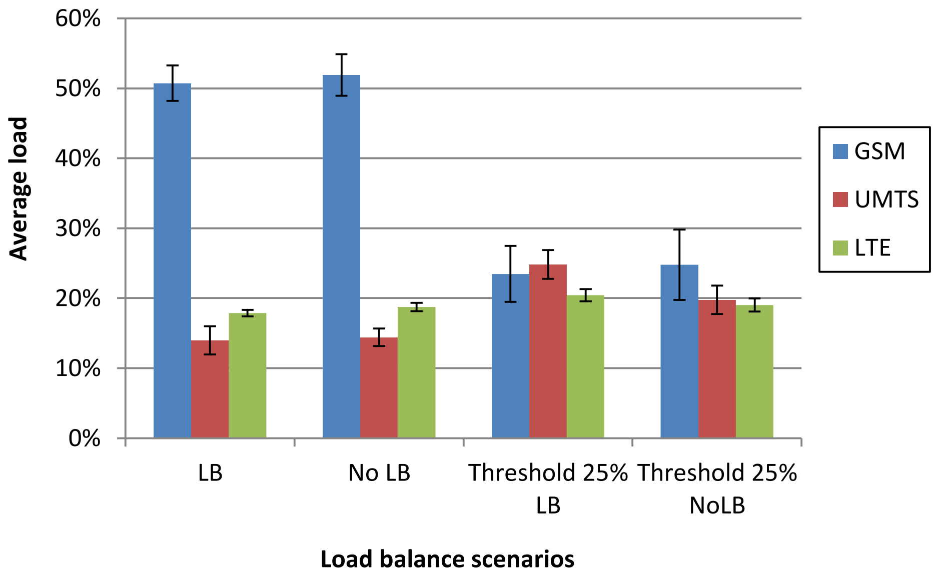


Figure 4-14 - LB model impact on load.

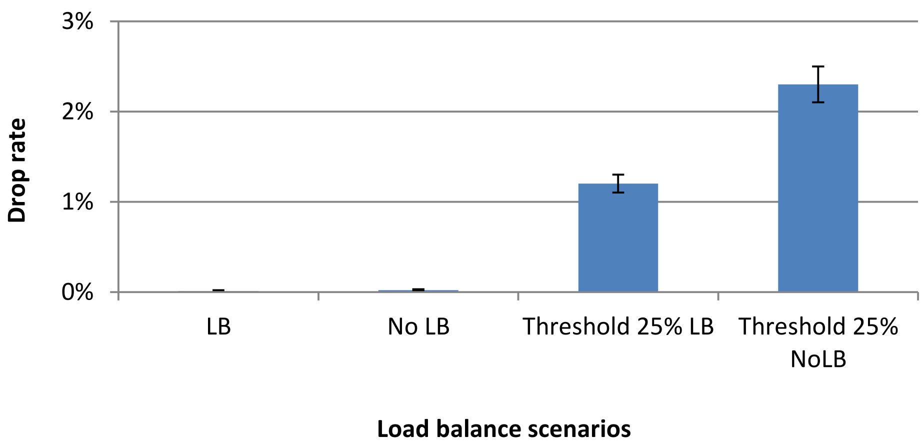


Figure 4-15 - LB model QoS – Drop rate

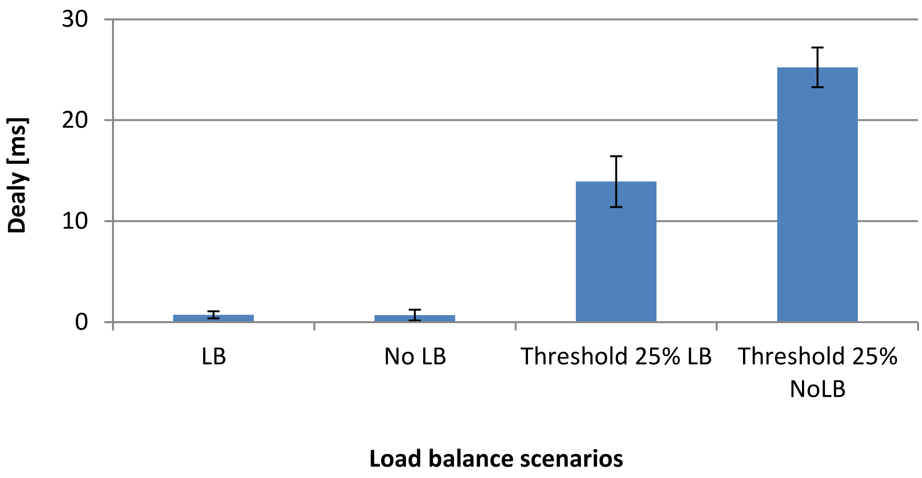


Figure 4-16 - LB model QoS – Delay

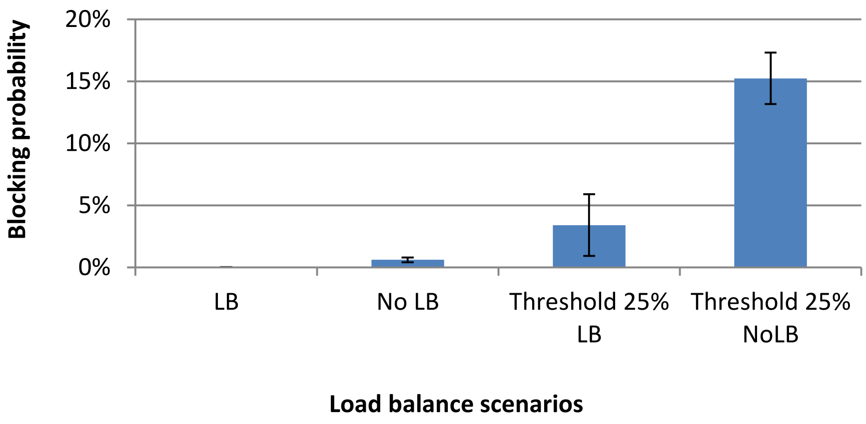


Figure 4-17 - LB model QoS – Blocking probability

From results, one may conclude that the model only has a notorious impact for an overloaded network. Only for the case where a 25% load threshold has been established, the model starts to react to the situation moving MTs from one BS to another, taking advantage of HHO and VHO to better distribute the traffic.

Figure 4-14 shows how the LB routine increases the average load in the network for all RATs, while increasing the number of users being served and providing a slight improvement in the analysed QoS parameters.

Chapter 5

Conclusions

The present chapter outlines the main conclusions of this master thesis. It reviews the thesis theme and purpose, thereafter highlights the most relevant results and ends with suggestions for future work to be developed within the area of LB in HNs.

Roughly every 10 years a great technology shift occurs in the wireless communications world introducing new systems and ways of communication. In Portugal, mobile devices penetration now exceeds more than one device per person, and the trend is to find many radio technologies working simultaneously in such devices like GSM/GPRS, UMTS/HSDPA, LTE, WiFi and Bluetooth. This wireless world is a heterogeneous world and operators exploit at their best the functionalities any given technology provides. GSM with its wide spread coverage and many years of settlement provides all the basic needs for reliable voice calls in almost any part of the country. UMTS has been going on for more than 10 years now, and is a reasonable system in terms of coverage, serving current traffic demands, although the appearance of smartphones and new cloud services are pushing demand to higher bitrates, only feasible with new technologies, like LTE.

With all this radio access technologies coexisting in the same space and at the same time, usually owned by the same telecom operator, naturally, the trend is to integrate the communication among all the systems in the most efficient way, transparently to the end user. This thesis exploits this need for networks' efficiency through load balancing in heterogeneous scenarios with GSM, UMTS and LTE. Henceforth, its main objective consisted of researching a technique to measure load in a heterogeneous scenario and to develop a model capable of balancing load among GSM, UMTS and LTE.

In Chapter 1, one gives a brief historical overview of the 3 systems under research and presents the thesis' theme and scope.

The complexity and heterogeneity of these systems required some study previous to model development, thereby, in Chapter 2, one presents the main characteristics for each system regarding their network architecture differences and radio access technologies. Furthermore, this chapter also presents the main steps involved in a handover procedure, essential for load balancing, as well as the state of the art focused in the thesis' scope.

Chapter 3 presents the approach to achieve the objective of the thesis, which can be split into two parts. At first, one had to find a common framework to compute load, regardless of the system, i.e., a given service may be overloading a UMTS base station, but its load may be better allocated to an LTE one, due to the different radio resources allocation scheme. Hence, some radio resources units were established to measure the impact of services in the base stations' load. Section 3.1 presents all the mathematical models to compute load, to measure the radio resource units per system, and to calculate the required resources for any given service. The second part of Chapter 3 consists of the algorithms and simulation tool used to implement the previously developed models. The simulation tool, SimCell, was developed upon previous versions from Serrador [Serr12] and Venes [Vene09]. A lot of effort was put in the simulator development to add two new systems (GSM and LTE), as well as making an overall overview on the main functionalities of the simulator to cope with these new systems. In the verge of the Internet of Things era, it would be unwise not to consider the impact of some new services already starting to be deployed; therefore, M2M services were also implemented and added to the existing ones. After its development, it required an intensive debugging and assessment phase due to the systems complexity and structural changes made.

The simulator is now capable of generating traffic for GSM, UMTS and LTE where users can seamlessly perform vertical handovers. The simulator has traffic models for 10 types of services, namely, voice calls, web browsing, video calls, e-mail, music streaming, file sharing, smart meters, eHealth, domotics and surveillance.

Due to the master thesis' scope, the modules regarding WiFi and WiMAX from previous versions were not further developed, in order to comply with same load balancing routines created for GSM, UMTS and LTE, and were left out of the simulator's development. Nonetheless, it will be possible, without much effort, to upgrade the simulator to be fully operational for GSM/GPRS, UMTS/HSDPA, LTE, WiFi, and WiMAX.

The results obtained in the present work are shown in Chapter 4, and their analysis shown the influence of many parameters in load within a heterogeneous scenario. Results show that the increase of the number of users in the scenario has a linear influence in the load state of base stations.

From simulation, one concludes that changing LTE's bandwidth has the biggest impact on the load state. By reducing the bandwidth from 10 MHz to 5 MHz, the load in the LTE base station increases by 12.3%; the theoretical analysis supports this idea, showing how a low bandwidth allocation for LTE leads to higher load states and fewer users allowed for connection.

Chapter 4 also presents the results and analysis of the implemented models. For this purpose, the simulation scenario considers a uniform distribution up to 9 000 users in a urban environment with two sites, both having base stations of the three types: GSM, UMTS and LTE. The results show, at first, little or no impact in the usage of the load balancing model, due to the abundance of RRUs in the scenario. The developed model is useful only in overloading situations.

In order to increase load in the scenario, one could increase the density of users and traffic demands, which would require a huge computational effort. Due to time constraints, such heavy load scenarios could not be tested, thereby, one was forced to lower the load threshold for all base stations to infer on the load balance algorithm.

After decreasing the load threshold from 70% to 25% the LB algorithm has shown some improvement in the overall QoS parameters analysed. The average load for all RATs has increased, while increasing the average number of users connected per second. The main results in terms of QoS are the reduction of drop rate calls from 2.3% to 1.2% and the reduction of blocking probability from 15.2% to 3.4% measured for the same scenario.

Note that such load threshold would not be realist in a real live network for the scenario tested, but it is a valid way to test the algorithm performance and to recreate a heavy load scenario like a music festival event.

It is important to take into account that, despite the effort to recreate a real wireless communication system, the world is far too complex, and many simplifications had to be introduced. The simulator performs most of the functionalities a radio management systems demands, but it does not strictly follow the existing communication protocols.

Despite all simplifications, the obtained results provide some insights and basic notions on how to deal with the load balancing challenge.

As suggestions for future work regarding the topic, it might be interesting to add some cost benefits on implementing this solution. The first version of Serrador's simulator considered such cost functions and it would have been an interesting subject to continue, this time, applied with the newly developed LTE and GSM modules.

Undoubtedly, much more could have been accomplished, and the present thesis is no more than a one-year effort where, above all, the author had the chance to learn more about the telecommunications' world and to have a first glance on what it takes to research and write within an academic context.

Annex A

COST 231 Walfish-Ikegami Model

This annex describes the propagation model used for the calculation of the path loss between transmitter and receiver.

The path loss between the emitted signal by a BS to a MT can be approximated by the COST 231 Walfish-Ikegami model, usually applied for urban and suburban scenarios for distances shorter than 5 km, whose input parameters, represented in Figure A-1, are the following ([Corr13]):

- h_b : height of BS antenna from ground;
- H_B : buildings height;
- h_m : MT height;
- w_s : street width;
- f : frequency;
- d : distance between BS and MT;
- w_B : building separation;
- ϕ : angle of incidence of the signal in the buildings.

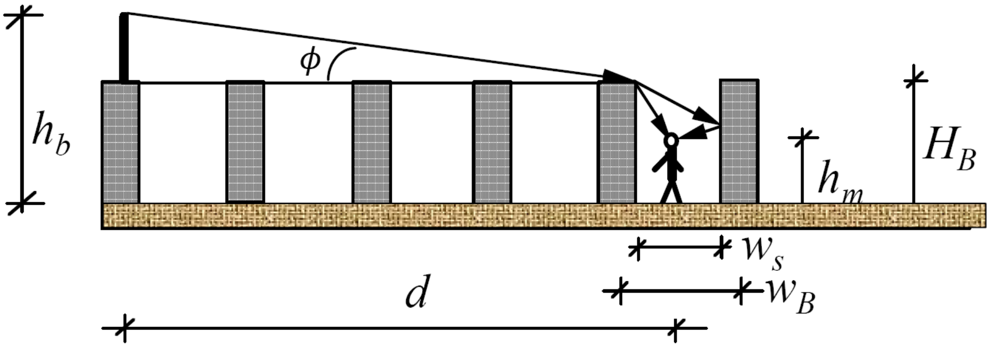


Figure A-1 - COST 231 Walfish-Ikegami model parameters (adapted from [Corr13]).

For Line-of-Sight (LoS) propagation in a street, and $d > 0.02$ km, path loss is given by:

$$L_{p[\mathrm{dB}]} = 42.6 + 26\log(d_{[\mathrm{km}]}) + 20\log(f_{[\mathrm{MHz}]}) \tag{A.1}$$

For all other cases, path loss is defined as:

$$L_{p[\mathrm{dB}]} = \begin{cases} L_{0[\mathrm{dB}]} + L_{rt[\mathrm{dB}]} + L_{rm[\mathrm{dB}]} & , L_{rt} + L_{rm} > 0 \\ L_{0[\mathrm{dB}]} & , L_{rt} + L_{rm} \leq 0 \end{cases} \tag{A.2}$$

where:

- L_0 : free space propagation path loss;
- L_{rt} : attenuation due to propagation from BS to the last rooftop;
- L_{rm} : attenuation due to diffraction from the last rooftop to the MT.

Being the path loss experienced in free space propagation given by:

$$L_{0[\mathrm{dB}]} = 32.44 + 20\log(d_{[\mathrm{km}]}) + 20\log(f_{[\mathrm{MHz}]}) \tag{A.3}$$

The propagation from the BS to the last rooftop experiences the following loss:

$$L_{rt[\mathrm{dB}]} = L_{bsh[\mathrm{dB}]} + k_a + k_d\log(d_{[\mathrm{km}]}) + k_f\log(f_{[\mathrm{MHz}]}) - 9\log\left(w_{B[\mathrm{m}]}\right) \tag{A.4}$$

where:

- $L_{bsh_{[dB]}} = \begin{cases} -18\log\left(h_{b_{[\mathrm{m}]}} - H_{B_{[\mathrm{m}]}} + 1\right) & , h_b > H_B \\ 0 & , h_b \leq H_B \end{cases}$

- $k_a = \begin{cases} 54 & , h_b > H_B \\ 54 - 0.8\left(h_{b_{[\mathrm{m}]}} - H_{B_{[\mathrm{m}]}}\right) & , h_b \leq H_B \wedge d \geq 0.5\ \mathrm{km} \\ 54 - 1.6\left(h_{b_{[\mathrm{m}]}} - H_{B_{[\mathrm{m}]}}\right) d_{[\mathrm{km}]} & , h_b \leq H_B \wedge d < 0.5\ \mathrm{km} \end{cases}$

- $k_d = \begin{cases} 18 & , h_b > H_B \\ 18 - 15\frac{h_{b[m]} - H_{B[m]}}{H_{B[m]}} & , h_b \leq H_B \end{cases}$

- $k_f = \begin{cases} -4 + 0.7\left(\frac{f_{[\mathrm{MHz}]}}{925} - 1\right) & , \text{urban and suburban scenarios} \\ -4 + 1.5\left(\frac{f_{[\mathrm{MHz}]}}{925} - 1\right) & , \text{dense urban scenarios} \end{cases}$

Finally, the loss due to diffraction from the last rooftop to the MT is given by:

$$L_{rm_{[\mathrm{dB}]}} = -16.9 - 10\log\left(w_{s_{[\mathrm{m}]}}\right) + 10\log\left(f_{[\mathrm{MHz}]}\right) + 20\log\left(H_{B_{[\mathrm{m}]}} - h_{m_{[\mathrm{m}]}}\right) + L_{ori_{[\mathrm{dB}]}} \tag{A.5}$$

with:

- $L_{ori_{[dB]}} = \begin{cases} -10.0 + 0.354\phi_{[°]} & , 0° < \phi < 35° \\ 2.5 + 0.075\left(\phi_{[°]} - 35\right) & , 35° < \phi < 55° \\ 4.0 - 0.114\left(\phi_{[°]} - 55\right) & , 55° \leq \phi \leq 90° \end{cases}$

The validity range for some parameters of this model imposes that:

- f ∈ [800; 2 000] MHz
- d ∈ [0.02; 5] km
- h_b ∈ [4; 50] m
- h_m ∈ [1; 3] m

The presented frequency range does not contain all the frequency bands studied in this thesis, therefore, one should consider higher relative errors than expected. The standard deviation of the model takes values in [4; 7] dB, and the error increases when h_b decreases relative to H_B.

In the absence of specific values, the following are recommended [Corr13]:

- w_B ∈ [20; 50] m
- $w_s = w_B/2$
- $\phi = 90°$
- $H_{B_{[m]}} = 3 \times (\#floors) + H_{roof}$
- $H_{roof_{[m]}} = \begin{cases} 3, pitched \\ 0, flat \end{cases}$

Annex B

Traffic Source Models

This annex presents the models used to simulate different traffic sources of the different services studied in the thesis.

B.1 Voice Models

For circuit-switched-based voice services, a 4-state model is used [VaRF99], which is based on measurement and includes not only the ON-OFF behaviour but also the effect of the voice encoder, compression device and air interface.

The model described in [VaRF99] and [RaMe01] defines four-states: when the source is in state $k1$ it generates packets of size S_{k1} each 10 ms, for a burst duration of τ_{k1}, $k1 = 1, \ldots, 4$. In the long time average, the probability that a packet is of size S_{k1} is quantified as P_{k1}. The burst duration τ_{k1} is modelled as a random variable; with mean value m_{k1}, with the Weibull Probability Density Function (PDF) as follows:

$$f(x) = \beta_{k1} \cdot \lambda_{k1} \cdot (x \cdot \lambda_{k1})^{\beta_{k1}-1} \cdot e^{-(x \cdot \lambda_{k1})^{\beta_{k1}}} \tag{B.1}$$

where:

- $1/\lambda_{k1}$ is the scale parameter;
- β_{k1} is the shape parameter;

and, both taking the values defined in Table B-1.

Table B-1 - Voice Source Model Parameters (partial extracted from [VaRF99]).

State $k1$	Packet size S_{k1} [Bytes]	Measured Probability P_{k1}	Measured mean burst duration m_{k1} [packet]	Weibull parameter λ_{k1}	Weibull parameter β_{k1}
1	2	0.5978	29.8	0.03	0.75
2	3	0.0723	2.5	0.45	0.80
3	10	0.0388	1.8	0.80	0.70
4	22	0.2911	38.8	0.05	0.90

After a $k1$ state, a new state is selected with probability Q_{k1}, which is defined as [RaMe01]:

$$Q_{k1} = \frac{\frac{P_{k1}}{m_{k1}}}{\sum_{j=1}^{4} \frac{P_j}{m_j}} \tag{B.2}$$

The voice calls generation process follows a Poisson process [Yaco93], and the duration of the calls is determined according to an exponential distribution.

VoIP services, in turn, present typically a symmetric or quasi-symmetric nature and require small end-to-end transmission delays. According to [Agui03], VoIP can be characterised through a traditional ON-OFF behaviour, in which sequences of speech-bursts are intercalated with silent bursts. Thus, a

VoIP transmission can be modelled as a Markov model with two states of "silence" and "talk": when in "silence", no packets are generated, and when in "talk", packets are generated at a constant rate. Particularly IP packets carrying the speech information are transmitted. Both activity and silent periods are generated by an exponential distributed random variable with mean values t_{ON} and t_{OFF}, respectively.

The payload size of the IP packets carrying speech bursts depends on the considered speech codec and the packet rate. Typical VoIP codecs are G711, G732.1 and G729.A, all of these with their specific frame duration and frame sizes, Table B–2.

Table B-2 – Typical VoIP codecs (extracted from [Nune02]).

Codec	Frame Duration [ms]	Frame Size [bytes]	Bitrate [kbps]
G711	10	80	64.0
G723.1	30	24	6.4
G729.A	20	20	8.0

As VoIP uses UDP (User Data Protocol) and RTP (Real Time Protocol) at the transport layer, the size of a full IPv6 (Internet Protocol version 6) header together with a RTP/UDP header is 60 bytes, and 40 bytes if IPv4 (Internet Protocol version 4) is used instead. As the size of a typical voice packet is 20 bytes if G729.A is used, the RTP/UDP/IP overhead figures illustrate the typical problem of the header overhead in VoIP: in this case, instead of an 8 kbps bitrate, a final bitrate of 32 kbps case IPv6 was in use would be generated (24 kbps if IPv4 is used instead). When operating in a bandwidth limited system such as WiMAX, it is important to use the radio band as effectively as possible, and header overhead up to 60 bytes can seriously degrade the spectral efficiency of a VoIP service over such link. Without header compression, two-thirds of the transmission would be just headers. In order to handle this purpose, protocols such as "RObust Header Compression" have been developed to tackle this problem [IETF01]. According to [Agui03], one can assume that header bytes can be compressed to 8 bytes. Additionally, and following European Telecommunications Standards Institute (ETSI) recommendations [ETSI98], speech calls should be generated according to a Poisson process, with mean call duration of 120 s. The resulting VoIP modelling is summarised in Table B–3.

Table B-3 - Modelling of VoIP Traffic.

Activity Factor [%]	50
Mean Active Phase, t_{ON} [s]	3
Mean Silent Phase, t_{OFF} [s]	3
Payload of IP Packets [bytes]	20
IP Overhead [bytes]	8
Transmission Time Interval [ms]	20
Mean Call Duration [s]	120

B.2 Video Model

In [ChRe98], a source model for variable bitrate video traffic model based on a finite-state Markov chain is presented, being demonstrated that it accurately models a one- and two-layers video. This model assumes two types of video frames generation, the *I* and *P* frames. *I* is driven by scene changes, depends on the video source, and as such can be considered independent of the encoder dynamics. However, in the time period characterisation by a sequence of *P* frames, one may expect that there are no significant changes in the information in successive frames. Therefore, the bit rates characterisation of successive *P* frames can be expected to be correlated, or more generally clustered around an average value. That fact allows a technique that uses correlation between successive frame bit rates to identify the *I* and *P* frames in the data. The frame rate ranges from 24 to 30 frames per second.

The *I* frame statistic is modelled by a Gaussian distribution function, which average and variance is adjusted as a function of measured data. *P* frames may be quite different, because they depend of video changes. In order to model these frames, a mechanism of $k1$ states was created, each state with its own average and variance.

The corresponding $k1$ state group to frame *P*, together with the only state that characterises frame *I*, typify the $K1$+1 Markov states that are represented by a probabilistic transitions matrix P. In this matrix, 90 % of the total of probabilistic transitions from one state to another are concentrated in $\{P_{ii-1}$, P_{ii}, $P_{ii+1}\}$ $\forall$ *I* =1,…, $k1$. Considering *i = n* any state in the Markov chain and *j = n-1* state.

The transmission speed of an *i* frame is given by [RaMe01], [ChRe98]:

$$\begin{cases} R_i(n) = m_i(1-\alpha_i) + \alpha_i R_j(n-1) + g_i(n) & i, j = 1, \ldots, k(P - frames) \\ R_{k+1}(n) = m_k + g_{k+1}(n) & (I - frames) \end{cases} \quad \text{(B.3)}$$

where:

- m_i is the estimated average value of transmission speed (*i* state)
- α_i is a coefficient value that models the autoregressive process, which value is adjusted empirically (based on measurements over *i* state in P frame)
- $g_i(n)$ is a Gaussian random variable, different for each state (average 0 and variance given by measurement data in *i* state)

B.3 Non-Conversational Applications

Non-real time applications typically present an asymmetrical nature, as they refer mostly to specific requests for information done by end users to remote machines. The most known applications are Web browsing, FTP and E-mail.

Several models are studied and proposed to characterise Web browsing but as the present work does not intend to focus specifically on traffic models, modelling of these applications is based on [ETSI98]. Figure A-1 illustrates a typical Internet surfing session, which consists of a sequence of packet calls. During a packet call several packets may be generated, which means that the packet call is composed of a bursty sequence of packets, Figure B-1.

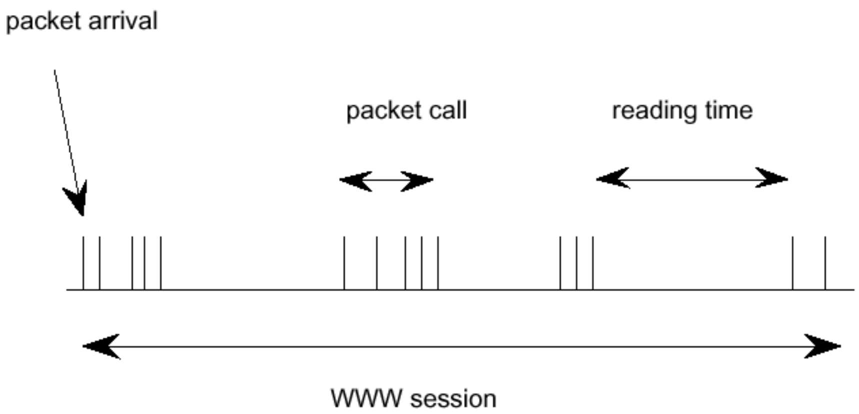


Figure B-1 - Typical WWW session (adapted from [ETSI98]).

A packet service session is modelled as a Poisson process. It typically contains one or several packet calls, depending on the application. In a WWW browsing session for instance, a packet call corresponds to the downloading of a web page, and after the document has entirely arrived to the terminal, the user takes some time for analysing the information, which is often called the reading time. On the other hand, in a FTP session, it is likely that the session contains only one packet call.

In order to fully characterise a packet session, the following parameters must be modelled [ETSI98]:

- The number of packet call requests per session, N_{pc}: This is a geometrically distributed random variable with mean μN_{pc}
- The reading time between two consecutive packet call requests in a session, D_{pc}: This is a geometrically distributed random variable with a mean μD_{pc}. Note that the reading time starts when the last packet of the packet call is completely received by the user. The reading time ends when the user makes a request for the next packet call.
- The number of packets within a packet call, N_d: Although different statistical distributions can be used to generate the number of packets, it is assumed that N_d can be a geometrically distributed random variable, with mean μN_d.

- The time interval between two consecutive packets inside a packed call, D_d: This is a geometrically distributed random variable with a mean μD_d. Naturally, if there is only one packet in a packet call, this is not needed.
- The Packet size, Ps: The packet size distribution model is based on Pareto distribution that suits best for the traffic case under study; Pareto distribution with cut-off is used.

B.4 Streaming Model

Audio and Video Streaming services typically present an asymmetrical nature, as they refer mostly to the download of large files. Contrary to VoIP, Streaming is more flexible to end-to-end transmission delays and to its variations, although the implementation of buffering is required to accommodate those fluctuations. Although Video traffic can be transported either with a constant bitrate or with a variable bitrate, variable bitrate is the one expected to be the prime example, as it has several potential advantages over constant bitrate, particularly the possibility for implementing statistical multiplexing, allowing for improved channel allocation. IEEE has not standardised a specific codec for Video Streaming yet. On the contrary, 3GPP has specified the use of both MPEG-4 and H.263 codecs for Video Streaming services [3GPP02].

For simplification reasons, the present work uses of the non-conversational applications model presented in A.3 also for Streaming service, although with different parameters, namely the mean values μN_{pc}, μN_d and packet size (parameters of Pareto distribution).

Annex C

Mobility Models

This annexes provides information on two mobility models used in this thesis to mimic the movement of real MTs. Section C.1 describes the Random Walk Mobility Model used to characterise the MT displacement, whereas for the MT's speed, it has been implemented Triangular Distribution Model for a more realist behaviour, shown in section C.2.

C.1 Random Walk Mobility Model

The Random Walk Mobility Model [CaBD02], Figure C–11, was developed to mimic the heretic behaviour of MTs giving a memory-less mobility patterns, as each step is calculated without any information of the previous one. At regular time intervals, both angular direction (uniformly distributed in [0º, 360º]) and speed of MTs are updated. MTs bounce at the border of the simulation area, in such a way that they can never roam outside this area. If the MT's speed or direction is updated frequently (short time intervals or distances), then it does not wander far off from the starting position. This effect can be useful when, e.g., one is simulating a semi-static network. If one wants to simulate more dynamic networks, larger values for the time interval or distance travelled should be used.

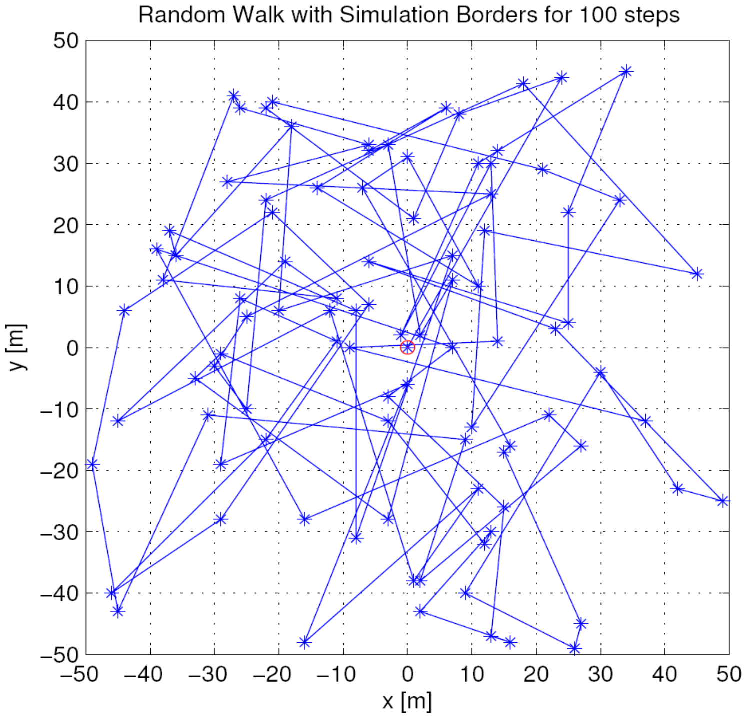


Figure C-1 - Random Walk Mobility pattern (extracted from [CaBD02]).

C.2 Triangular Distribution Mobility Model

The model presented in [ChLu95] considers a triangular distribution for speed with five different mobility types, as shown in Table C–1 with average velocity given by:

$$\mathrm{V}_{av} = (V_{max} + V_{min})/2 \tag{C.1}$$

and deviation:

$$\Delta = (V_{max} - V_{min})/2 \tag{C.2}$$

Table C-1 - Mobility type speed characteristics (adapted from [ChLu95]).

Mobility Type	V_{av} **[m/s]**	**Δ [m/s]**
Static	0	0
Pedestrian	1	1
Urban	10	10
Main Roads	15	15
Highways	22.5	12.5

The density function of the Triangular Distribution Model, shown in Figure C-2, is given by the following expression:

$$\mathrm{f}(\nu) = \begin{cases} \frac{1}{\Delta^2}[\nu - (V_{av} - \Delta)] & \text{, if } V_{av} - \Delta \leq \nu \leq V_{av} \\ -\frac{1}{\Delta^2}[\nu - (V_{av} - \Delta)] & \text{, if } V_{av} \leq \nu \leq V_{av} + \Delta \\ 0 & \text{, otherwise} \end{cases} \tag{C.3}$$

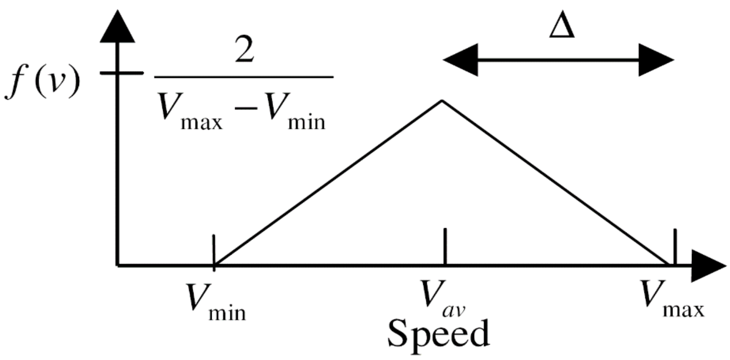


Table C-2 - Velocity probability density function (extracted from [ChLu95]).

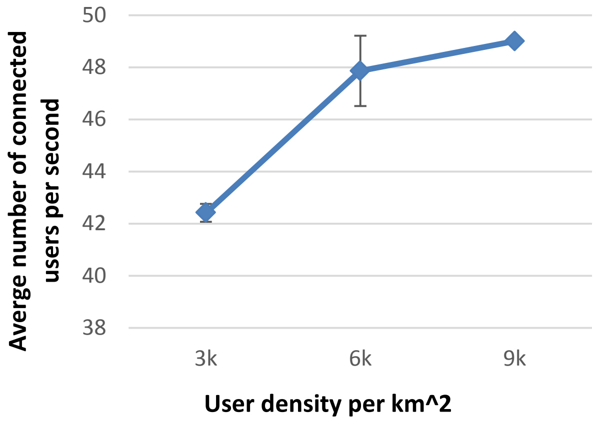
Annex D

Simulator's assessment results for GSM and UMTS RANs

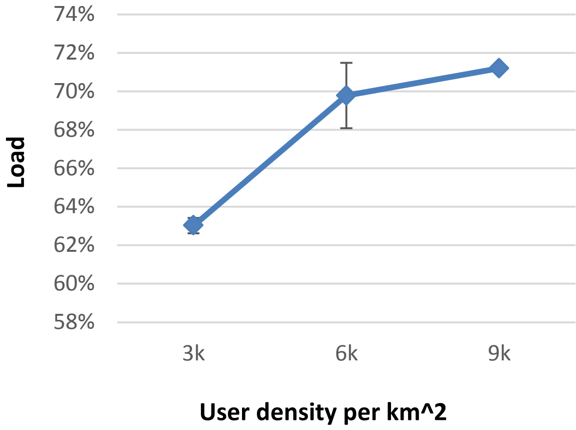
This annex has the plots and results obtained in the assessment of the GSM and UMTS RANs implemented in the SimCell simulator.

The present annex presents the graphical results from the simulator assessment and the analysis in detail can be found at Chapter 3, section 3.2.3.

In the case of a single GSM in Figure D–1 (a), above 6 000 users/km^2, the system has surpassed the 70 % threshold load. Figure D–1 plots the assessment analysis in the number of connected users and load state.



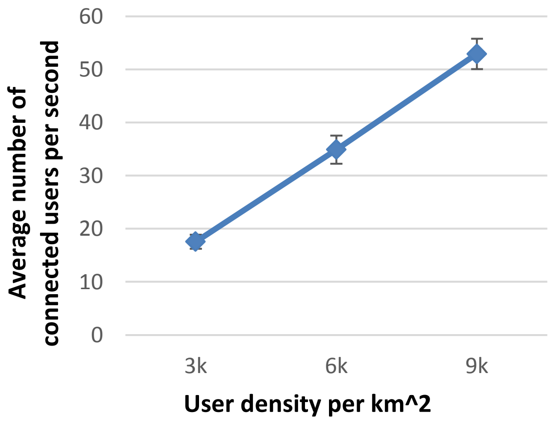
(a) Number of connected users



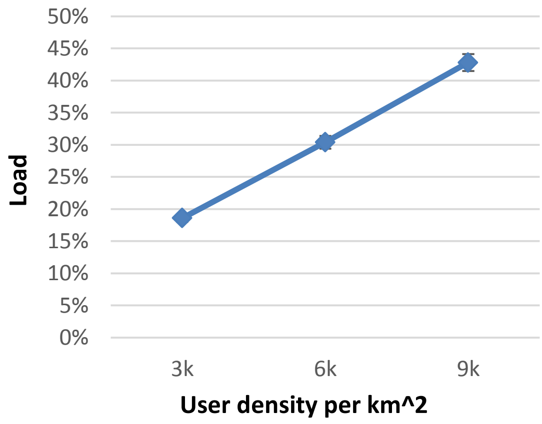
(b) Load

Figure D-1 - GSM Assessment.

The same analysis has been made for UMTS, registered in Figure D–2.



(a) Number of connected users



(b) Load

Figure D-2 - UMTS Assessment.

The same load analysis on the load evolution for GSM and UMTS are represented below, in Figure D-3 and Figure D–4, respectively. All three RATS had registered an initial convergence state as well as a fading down load profile at the end of the simulation. GSM in particular, has reached the load threshold of 70% for the assessment scenario, and for that reason presents a slight different load evolution profile comparing to the LTE and UMTS profile.

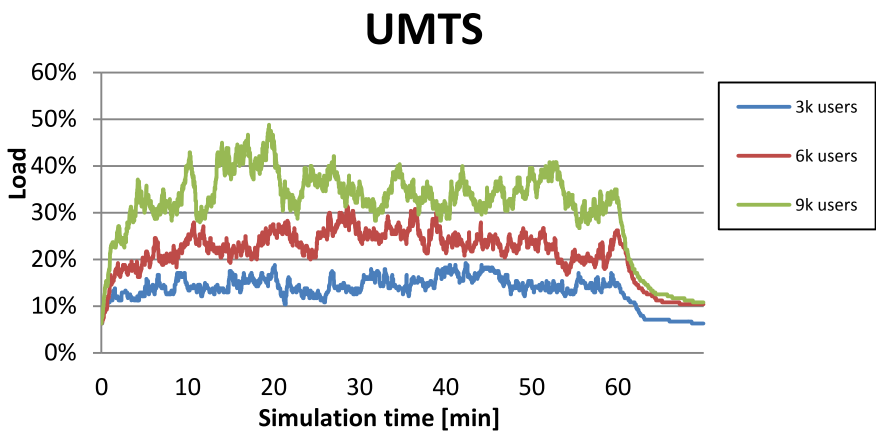


Figure D-3 - UMTS Assessment - Load time evolution.

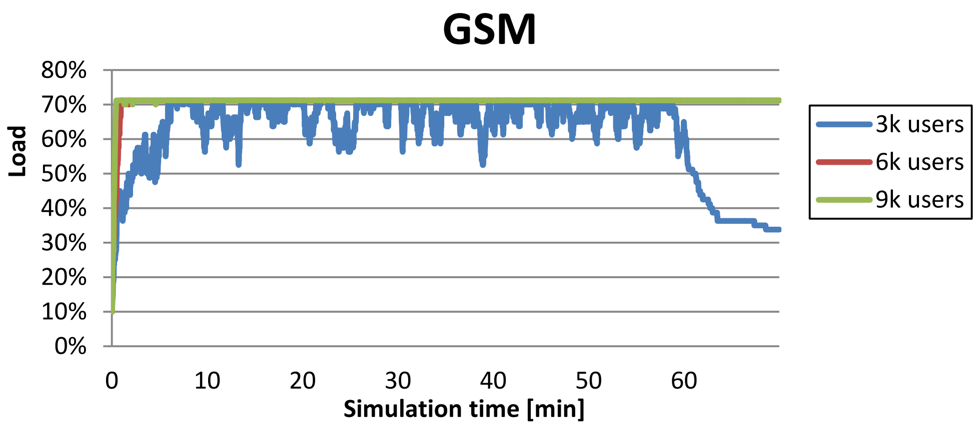


Figure D-4 - GSM Assessment - Load time evolution.

References

[3GPP02] 3GPP, *Transparent end-to-end packet switched streaming service (PSS); Protocols and codecs (Release 4)*, 3GPP Technical Specification No. 26.234 Ver. 4.3.0, Valbonne, France, Mar. 2002. (www.3gpp.org).

[3GPP08] 3GPP, "UTRA-UTRAN Long Term Evolution (LTE) and 3GPP System Architecture Evolution (SAE)", White Paper, 2008, Dec. 2013, (ftp://ftp.3gpp.org/Inbox/2008_web_files/LTA_Paper.pdf).

[Agil09] Agilent Technologies, "Agilent 3GPP Long Term Evolution: System Overview, Product Development and Test Challenges", Application Note, USA, 2009. (http://cp.literature.agilent.com/litweb/pdf/5989-8139EN.pdf).

[ANAC12] ANACOM, "Relatório final do Leilão", Dec. 2013, (http://www.anacom.pt/streaming/Relat_final_leilao_Jan2012.pdf?contentId=1112754&field=ATTACHED_FILE).

[BeJu03] Bernhard,U., Jugl,E., Mueckenheim,J., Pampel,H., and Soellner,M., "Intelligent Management of Radio Resources in UMTS Access Networks", *The Bell Labs Technical Journal*, Vol.7, 2003, pp. 109-126.

[BiHJ03] Bing,H., He,C. and Jiang,L., "Performance Analysis of Vertical Handover in UMTS-WLAN Integrated Network", in *Proc. of PIMRC'03 – 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Beijing, China, Sep. 2003.

[Carr11] Carreira,P., *Data Rate Performance Gains in UMTS Evolution to LTE*, M.Sc. Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2011.

[CaBD02] Camp,T., Boleng,F. and Davies,V., "A Survey of Mobility Models for Ad Hoc Network Research", *Wireless Communication & Mobile Computing*, Vol. 2, No. 5, 2002, pp 483-502.

[CCOJ05] Chebbine,S., Chebbine,M.T., Obaid,A. and Johnston,R., "Framework Architecture for Internet Content Adaptation System and Vertical Handover Management on 4G Networks" in *Proc. of WTS'05 – 4th Annual Wireless Telecommunications Symposium*, Pomona, CA, USA, Apr. 2005.

[ChJa89] Chiu,D. and Jain,R., "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks", *Computer Networks and ISDN systems*, Vol.17, No.1, 1989, pp. 1–14.

[ChLu95] Chlebus,E. and Ludwin,W., "Is handoff traffic really Poissonean?", in *Proc. of*

ICUPC'95 – 4th IEEE International Conference on Universal Personal Communications, Tokyo, Japan, Nov. 1995.

[ChRe98] Chandra,K. and Reibman,A.R., "Modeling One and Two Layer Variable Bit Rate Video", *IEEE/ACM Transaction on Networking*, Vol. 7, No. 3, 1998, pp. 348-413.

[CISC13] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018", White Paper, Feb.2014, (http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf).

[Corr13] Correia,L.M., "Mobile Communications Systems – Lecture Notes", Instituto Superior Técnico, Lisbon, Portugal, 2013.

[DaPS11] Dahlman,E., Parkvall,S., Sköld,J., *4G LTE/LTE-Advanced for Mobile Broadband*, Academic Press, Oxford, UK, 2011.

[Duar08] Duarte,S., *Analysis of Technologies for Long Term Evolution in UMTS*, M.Sc. Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2008.

[EART11] Energy Aware Radio and network tecHnologies (ICT-EARTH), Jan. 2011 (https://www.ict-earth.eu)

[Eric03] Ericsson, Handover between WCDMA and GSM, 2003 (http://www.ericsson.com/ph/res/thecompany/docs/publications/ericsson_review/2003/2003011.pdf).

[ETSI98] ETSI, *Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS*, ETSI Technical report No. TR 101 112 Ver. 3.2.0, France, Apr. 1998.

[FGSP05] Ferrús,R, Gelonch,A. Sallent,O., Pérez-Romero,J., Nafisi,N. and Dohler,M., "Vertical Handover Support in Coordinated Heterogeneous Radio Access Networks", in *Proc. of IST-MWCS'2005 – 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005.

[HoTo04] Holma,H. and Toskala,A., *WCDMA for UMTS – Radio Access For Third Generation Mobile Communications*, John Wiley & Sons, Chichester, UK, 2004.

[HoTo07] Holma,H. and Toskala,A., *WCDMA for UMTS – HSPA Evolution and LTE (4th ed.)*, John Wiley & Sons, Chichester, UK, 2007.

[HoTo09] Holma,H. and Toskala,A., *LTE for UMTS OFDMA and SC-FDMA Based Radio Access*, John Wiley & Sons, Chichester, UK, 2009.

[HoTo11] Holma,H. and Toskala,A., *LTE for UMTS: Evolution to LTE Advanced (2nd ed.)*, John Wiley & Sons, Chichester, UK, 2011.

[ICT13] ITU-D, ICT Facts and Figures 2013, Dec. 2013, (http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013.pdf).

[IETF01] IETF, Robust Header Compression (ROHC): Framework and four profiles, IETF RFC 3095, Center Oak Plaza, VA, USA, July 2001 (http://www.ietf.org/rfc.html).

[KYII10] Kanno,I., Yamazaki,K., Ikeda,Y. and Ishikawa,H., "Adaptive Energy Centric Radio Access Selection for Vertical Handover in Heterogeneous Networks", in *Proc. of WCN 10 – IEEE Wireless Communications and Networking Conference*, Sydney, Australia, Apr. 2010 (http://www.ieee-wcnc.org/2010).

[LeMa01] Lempiänien,J. and Manninen,M., *Radio Interface System Planning or GSM/GPRS/UMTS*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.

[Lobi10] Lobinger,A., Stefanski,S. Jansen,T. and Balan,I., *Load balancing in downlink LTE self-optimising networks*, VTC2010-Spring, Taipei, Taiwan, May. 2010.

[Moli11] Molisch,A. F, *Wireless Communications*, John Wiley & Sons, Chichester, UK, 2011.

[NSN10] Nokia Siemens Networks, "Mobile Broadband with HSPA and LTE – Capacity and Cost Aspects", White Paper, Last seen in Dec. 2013 (http://nsn.com/system/files/document/Mobile_broadband_A4_26041.pdf).

[NgAg11] Nguyen-Vuong,Q., Agoulmine,N., *"*Efficient Load Balacing Algorithm over Heterogeneous Wireless Packet Networks*"*, *REV Journal on Electronics and Communications*, Vol.1, No.1, Jan-Mar,2011.

[Nune02] Nunes,M.S., *Integrated Services Networks – 4th section – Voice over IP* (in Portuguese), AEIST, Lisbon, Portugal, 2002.

[Proa01] Proakis,J.G., *Digital Communications*, MacGraw Hill, New York, USA, 2001.

[RaMe01] Rábanos,J. and Mesquida,C., *Mobile Communication of Third Generation*, (In Spanish), Telefónica Móviles España, Madrid, Spain, 2001.

[Rono12] Ronoh,K., *Load Balancing in Heterogeneous LTE-A Networks*, M.Sc. Thesis, Linköping University, Norrköping, Sweden, 2012.

[Rumn09] Rumney,M., *LTE and the Evolution to 4G Wireless – Design and Measurement Challenges*, Agilent Technologies, Padstow, Cornwall, 2009.

[Rysa10] Rysavi, "Transition to 4G – 3GPP Broadband Evolution to IMT-Advanced", Research Report, Rysavi Research, Hood River, USA, Sep. 2010 (http://www.4gamericas.org/documents/3G_Americas_RysavyResearch_HSPA-LTE_Advanced_FINALv1.pdf).

[Saut11] Sauter,M., *From GSM to LTE – An Introduction to Mobile Networks and Mobile Broadband*, John Wiley & Sons, Chichester, UK, 2011.

[Serr02] Serrador,A., *Optimisation of cell radius in UMTS-FDD networks*, M.Sc. Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2002.

[Serr12] Serrador,A., *Joint Radio Resource Management in Heterogeneous Networks*, Ph.D. Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2012.

[SeTB11] Sesia,S., Toufik,I. and Baker,M., *LTE - The UMTS Long Term Evolution*, John Wiley & Sons, Chichester, UK, Second Edition, 2011.

[Silv12] Silva,D., *Energy efficient solutions in GSM/UMTS based on traffic profiling models*, M.Sc. Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2012.

[VaRF99] Valko,A.G., Racz,A. and Fodor,G., "Voice QoS in Third Generation Mobile Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 17, No. 1, Jan. 1999, pp. 109-123.

[VeAl04] Velayos,H., Aleo,V., and Karlsson,G., "Load Balancing in Overlapping Wireless LAN Cells," in *Proc. of ICC - IEEE Int. Conference on Communications*, Kista, Sweden, 2004.

[Vene09] Venes,J., *Common Radio Resources Management in Heterogeneous Networks, with UMTS, Wi-Fi and WiMAX*, M.Sc. Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2009.

[Walk02] Walke,B.H., *Mobile Radio Networks – Networking, Protocols and Traffic Performance*, John Wiley & Sons, Chichester, UK, 2002.

[ZdSc04] Zdarsky,F.A. and Schmitt,J.B., "Handover in Mobile Communication Networks: Who is in Control Anyway?", In *Proc.* of *EUROMICRO – 30th IEEE EUROMICRO Conference*, Rennes, France, Sep. 2004.