Capacity Management and Load Balancing for 3G Wireless Access Networks

Tiago Miguel de Almeida Pedro

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisors:  Prof. António José Castelo Branco Rodrigues
              Prof. Pedro Manuel de Almeida Carvalho Vieira

Examination Committee

Chairperson:  Prof. José Eduardo Charters Ribeiro da Cunha Sanguino
Supervisor:   Prof. António José Castelo Branco Rodrigues
Members of the Committee: Prof. Pedro Joaquim Amaro Sebastião
                        André Eduardo Ponciano Martins

May 2016
Acknowledgements

This thesis brings to a conclusion a very important stage of my life and represents the highlight of the previous five years of superior education. In order to accomplish this, the support of many people and institutions was indispensable. In the following paragraphs I will make a small tribute to all of those who made the development of this thesis possible.

Firstly, I would like to thank my parents for all the support, help and motivation they gave me throughout my education. Their teachings and guidance made me develop into a good professional and a good person. Moreover, I would like to thank all my family, whose love and caring were present along all my life.

Secondly, I would like to thank everyone who worked closely with me during the course of my thesis. To Prof. António Rodrigues and Prof. Pedro Vieira, my supervisors, I thank all the advices, guidance and support given throughout the development of this work. I would also like to thank Celfinet for giving me the chance to work in close collaboration with them, and particularly to Engineers Nuno Oliveira Silva, André Martins and Tiago Cunha, who provided crucial insight to the development of this thesis.

I would also like to extend my thanks to all the educational institutions that marked my academic life, Academia de Música de Santa Cecília, Instituto Superior Técnico and the Norwegian University of Science and Technology (NTNU). They gave me invaluable knowledge and also enabled my development as a person.

Last but not least, I want to thank my friends, particularly the ones from my childhood, my university and my Erasmus experience. They stood next to me in good and bad times, giving me the strength and motivation to keep going. Without them none of this would be possible.


**Resumo**

De forma a sobreviver num mercado extremamente competitivo, as operadoras de redes móveis têm de gerir os seus recursos da forma mais eficiente possível. Esta gestão torna-se ainda mais relevante quando o recurso em questão é a capacidade instalada nas estações base da rede. Esta tese tem como objectivo não só fornecer aos operadores uma ferramenta para prever o momento em que as suas estações base atingirão o seu limite de capacidade, utilizando um algoritmo de previsão de tráfego, mas também propor um método de prolongar a sua longevidade através da implementação de um algoritmo de balanceamento de carga. Este último algoritmo tira partido da existência de estações com co-localização de duas bandas de frequência em operação, sugerindo um *threshold* de admissão para a banda de frequência mais elevada, e dessa forma distribuindo o tráfego de forma mais eficiente pelas duas bandas. Os dados utilizados para *input* dos dois algoritmos são as estatísticas reais de tráfego, incluindo indicadores geo-posicionados, fornecidas por uma consultora portuguesa de telecomunicações. Ao longo do decorrer desta tese foi possível desenvolver um método semi-automático para a optimização da rede, contribuindo para o progresso das *Self-Organizing Networks* nacionais. Este trabalho foi desenvolvido em colaboração com uma empresa de consultoria de telecomunicações portuguesa, a Celfinet, que providenciou supervisão e orientação indispensáveis. Utilizando o método sugerido prevê-se que, ao fim de um ano de implementação, seja possível atingir poupanças na ordem dos 70% em equipamento rádio para expansões de capacidade de estações base.

Abstract

In order to survive in a highly competitive market, mobile network operators have to be as efficient as possible in managing their resources. This is particularly relevant in what concerns the capacity available at their sites. This thesis aims to give the operators a method to predict when their sites will reach their capacity limit, through the implementation of a forecast algorithm, and also a method to improve the longevity of sites through the use of a Load Balancing algorithm. This latter algorithm relies on the existence of sites with co-located frequency layers, by suggesting an admission threshold for the higher frequency band, thus more efficiently distributing traffic between the two bands. The input data used consisted of real traffic statistics, including geo-located indicators, provided by a Portuguese telecommunications consulting company. During the course of this thesis it was possible to develop an semi-automatic method for network optimization using real and geo-located data, thus making a contribute to the development of the national Self-Organizing Networks. This project was developed in collaboration with a Portuguese telecommunications consulting company, Celfinet, which provided valuable supervision and guidance. Using the suggested method it is predicted that, after a year of implementation, it is possible to achieve savings of about 70% in capacity expansions in the network.

Keywords: 3G Wireless Access Networks, Traffic Forecasting, Load Balancing, Capacity Management, Multi-layer Site Optimization, Self Organizing Networks, Geo-positioned Indicators.
Contents

Acknowledgements iii

Resumo v

Abstract vii

Contents ix

List of Figures xi

List of Tables xiii

Abbreviations xv

Symbols xix

1 Introduction 1
  1.1 Motivation .......................... 1
  1.2 Approach ............................. 2
  1.3 Structure ............................. 3

2 State of the Art 5
  2.1 3G Wireless Access Networks ........................................ 5
      2.1.1 Radio Interface ........................................ 6
      2.1.2 Radio Access Network Architecture ........................................ 10
      2.1.3 Frequency Bands ........................................ 13
      2.1.4 Handover ........................................ 14
      2.1.5 Transport Channels ........................................ 15
      2.1.6 Downlink Cell Capacity ........................................ 18
      2.1.7 Capacity Management in UMTS ........................................ 20
  2.2 Load Balancing ........................................ 21
      2.2.1 Capacity Driven Load Balancing ........................................ 21
      2.2.2 Dynamic Handover Algorithms ........................................ 24
  2.3 Traffic Analysis ........................................ 25
2.3.1 Erlang Traffic Models .................................................. 25
2.3.2 ON/OFF Model ......................................................... 27
2.3.3 Traffic Prediction Models ............................................. 28

3 Traffic Forecasting .......................................................... 31
  3.1 Overview ................................................................. 31
  3.2 Input Data ............................................................... 32
    3.2.1 RAN Topology File ............................................... 33
    3.2.2 Traffic Statistics Files ......................................... 34
    3.2.3 Data Parsing ..................................................... 36
  3.3 Weekly Site Characterization ........................................ 36
  3.4 Forecasting Algorithm ................................................. 38
    3.4.1 Fitting and Decision Stages .................................... 39
    3.4.2 Validation Stage ................................................ 40
    3.4.3 Results and Analysis .......................................... 42

4 Load Balancing ............................................................. 49
  4.1 Overview ............................................................... 49
  4.2 Input Data ............................................................. 50
  4.3 Imbalance Analysis .................................................. 51
  4.4 Load Balancing Algorithm .......................................... 54
    4.4.1 Parameter Estimation for the U900 Band ...................... 54
    4.4.2 Load Balancing Strategy ....................................... 55
    4.4.3 Results and Analysis .......................................... 57

5 Capacity Limit Prediction ................................................ 63
  5.1 Overview ............................................................. 63
  5.2 Input Data ............................................................. 63
  5.3 Improved Forecasting ................................................. 64
  5.4 Results and Analysis ................................................. 67

6 Conclusions ........................................................................ 73
  6.1 Summary ................................................................. 73
  6.2 Future Improvements .................................................. 74

A Blocking Probability Models ............................................ 81
List of Figures

2.1 Spreading and Despreading in DS-CDMA ......................................................... 8
2.2 DS-CDMA correlation receiver principle ......................................................... 8
2.3 High-level UMTS architecture ........................................................................... 10
2.4 Detailed UMTS architecture .............................................................................. 11
2.5 UTRAN architecture ......................................................................................... 11
2.6 Mapping of transport channels onto physical channels ....................................... 18
2.7 Example of inter-frequency load balancing ....................................................... 22
2.8 ON/OFF model ................................................................................................... 27

3.1 Mapping of sites according to their weekly usage .............................................. 37
3.2 Overview flowchart of the forecasting algorithm ............................................... 38
3.3 Splitting of input data into the 3 sets .................................................................... 39
3.4 Flowchart of the fitting and decision stages of the forecast algorithm ............... 41
3.5 Flowchart of the validation stage of the forecast algorithm ............................... 42
3.6 Example of a forecast used for validation .......................................................... 42
3.7 Example of a forecast used for validation, with weekly input data ..................... 43
3.8 Example of a forecast used for validation, with speech traffic as input data ........ 43
3.9 Example of a forecast used for validation, with HSPA traffic as input data ......... 44
3.10 Example of a forecast used for validation, with R99 data traffic as input data ..... 44
3.11 Example of a forecast used for validation, with user number as input data .......... 45

4.1 Imbalance factors for all sites ............................................................................. 52
4.2 Imbalance differential CDF for all Sites ............................................................ 54
4.3 Flowchart of the Load Balancing algorithm ....................................................... 58
4.4 Imbalance factors for all sites, before and after Load Balancing ....................... 58
4.5 Imbalance differential CDF after Load Balancing ............................................. 60
4.6 Coverage of Site 24 after Load Balancing ......................................................... 61
4.7 Coverage of Site 20 after Load Balancing .......................................................... 61

5.1 Example of input data used for Forecasting ...................................................... 65
5.2 Prediction error distribution for selected data input ........................................... 67
5.3 Prediction error distribution for selected data input ........................................... 67
5.4 Example of Forecast before Load Balancing .................................................... 68
5.5 Example of Forecast after Load Balancing ....................................................... 68
5.6 Capacity Limit complementary CDF ................................................................. 69
5.7 CEs needed for expansion of all sites over one year .......................................... 70
# List of Tables

2.1 Comparison between UMTS 900 MHz and UMTS 2100 MHz bands . . . . . . . . 14  
2.2 Assumptions in the calculation of estimated downlink throughput . . . . . . . . 19  
3.1 Characteristics of available input data for traffic forecasting . . . . . . . . . . . . 32  
3.2 Summary of available input data for traffic forecasting . . . . . . . . . . . . . . 35  
3.3 Site characterization according to weekly usage . . . . . . . . . . . . . . . . . . 37  
3.4 Average MAPE for the several input data types . . . . . . . . . . . . . . . . . . 45  
3.5 Average NMRSE for the several input data types . . . . . . . . . . . . . . . . . . 46  
3.6 Usage of each function in the forecast algorithm . . . . . . . . . . . . . . . . . . 47  
3.7 Prediction error associated with each function chosen by the forecast algorithm . 47  
4.1 Characteristics of available input data for load balancing . . . . . . . . . . . . . . 50  
4.2 Imbalance factors for selected sites . . . . . . . . . . . . . . . . . . . . . . . . . 53  
4.3 Summary of Load Balancing results . . . . . . . . . . . . . . . . . . . . . . . . . 59  
5.1 Prediction error grouped by fitting function chosen . . . . . . . . . . . . . . . . . 65  
5.2 Prediction error for improved Forecasting algorithm . . . . . . . . . . . . . . . . 66  
5.3 Capacity Limit results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71
Abbreviations

2G          Second Generation
3G          Third Generation
3GPP        3rd Generation Partnership Project
4G          Fourth Generation
ATM         Asynchronous Transfer Mode
ARIMA       Auto-Regressive Integrated Moving Average
ARMA        Auto-Regressive Moving Average
ARQ         Automatic Repeat Request
BCH         Broadcast Channel
CapEx       Capital Expenditure
CAT         Concentrated Adaptive Threshold
CCTrCh      Coded Composite Transport Channel
CDF         Cumulative Distribution Function
CDR         Call Drop Rate
CE          Channel Element
CN          Core Network
CPICH       Common Pilot Channel
CS          Circuit-Switched
CSSR        Call Setup Success Rate
CSV         Comma Separated Values
DAT         Distributed Adaptive Threshold
DCH         Dedicated Channel
DL          Downlink
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPCCH</td>
<td>Dedicated Physical Control Channel</td>
</tr>
<tr>
<td>DPDCH</td>
<td>Dedicated Physical Data Channel</td>
</tr>
<tr>
<td>DS-CDMA</td>
<td>Direct-Sequence Code Division Multiple Access</td>
</tr>
<tr>
<td>E-AGCH</td>
<td>E-DCH Absolute Grant Channel</td>
</tr>
<tr>
<td>E-DCH</td>
<td>Enhanced DCH</td>
</tr>
<tr>
<td>E-DPCCH</td>
<td>Enhanced DPCCH</td>
</tr>
<tr>
<td>E-DPDCH</td>
<td>Enhanced DPDCH</td>
</tr>
<tr>
<td>eFACH</td>
<td>Enhanced FACH</td>
</tr>
<tr>
<td>E-HICH</td>
<td>E-DCH HARQ Indicator Channel</td>
</tr>
<tr>
<td>E-RGCH</td>
<td>E-DCH Relative Grant Channel</td>
</tr>
<tr>
<td>EUL</td>
<td>Enhanced Uplink</td>
</tr>
<tr>
<td>FACH</td>
<td>Forward Access Channel</td>
</tr>
<tr>
<td>FARIMA</td>
<td>Fractional Auto-Regressive Integrated Moving Average</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency Division Duplex</td>
</tr>
<tr>
<td>FT</td>
<td>Fixed Threshold</td>
</tr>
<tr>
<td>GGSN</td>
<td>Gateway GPRS Support Node</td>
</tr>
<tr>
<td>GMSC</td>
<td>Gateway Mobile Services Switching Centre</td>
</tr>
<tr>
<td>GPRS</td>
<td>General Packet Radio System</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communications</td>
</tr>
<tr>
<td>HARQ</td>
<td>Hybrid ARQ</td>
</tr>
<tr>
<td>HSDPA</td>
<td>High-Speed Downlink Packet Access</td>
</tr>
<tr>
<td>HS-DPCCH</td>
<td>Uplink High-Speed Dedicated Physical Control Channel</td>
</tr>
<tr>
<td>HS-DSCCH</td>
<td>High-Speed Downlink Shared Channel</td>
</tr>
<tr>
<td>HS-SCCH</td>
<td>High-Speed Shared Control Channel</td>
</tr>
<tr>
<td>HSPA</td>
<td>High-Speed Packet Access</td>
</tr>
<tr>
<td>HSPA+</td>
<td>High-Speed Packet Access Evolution</td>
</tr>
<tr>
<td>HSUPA</td>
<td>High-Speed Uplink Packet Access</td>
</tr>
<tr>
<td>HLR</td>
<td>Home Location Register</td>
</tr>
<tr>
<td>IAF</td>
<td>Intra-frequency neighbour related</td>
</tr>
<tr>
<td>IEF</td>
<td>Inter-frequency neighbour related</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>LB</td>
<td>Load Balancing</td>
</tr>
<tr>
<td>LTE</td>
<td>Long-Term Evolution</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean Absolute Percentage Error</td>
</tr>
<tr>
<td>ME</td>
<td>Mobile Equipment</td>
</tr>
<tr>
<td>MGW</td>
<td>Media Gateway</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple Input Multiple Output</td>
</tr>
<tr>
<td>MSC</td>
<td>Mobile Services Switching Centre</td>
</tr>
<tr>
<td>NRMSE</td>
<td>Normalized Mean Squared Error</td>
</tr>
<tr>
<td>OpEx</td>
<td>Operational Expenditure</td>
</tr>
<tr>
<td>PCH</td>
<td>Paging Channel</td>
</tr>
<tr>
<td>PS</td>
<td>Packet-Switched</td>
</tr>
<tr>
<td>PLMN</td>
<td>Public Land Mobile Network</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>QPSK</td>
<td>Quadrature Phase Shift Key</td>
</tr>
<tr>
<td>RAB</td>
<td>Radio Resource Access Bearer</td>
</tr>
<tr>
<td>RACH</td>
<td>Random Access Channel</td>
</tr>
<tr>
<td>RAN</td>
<td>Radio Access Network</td>
</tr>
<tr>
<td>RNC</td>
<td>Radio Network Controller</td>
</tr>
<tr>
<td>RNS</td>
<td>Radio Network Sub-system</td>
</tr>
<tr>
<td>ROP</td>
<td>Report Output Period</td>
</tr>
<tr>
<td>RRC</td>
<td>Radio Resource Control</td>
</tr>
<tr>
<td>RRM</td>
<td>Radio Resource Management</td>
</tr>
<tr>
<td>RSCP</td>
<td>Received Signal Code Power</td>
</tr>
<tr>
<td>RTWP</td>
<td>Received Total Wideband Power</td>
</tr>
<tr>
<td>SC</td>
<td>Spreading Code</td>
</tr>
<tr>
<td>SF</td>
<td>Spreading Factor</td>
</tr>
<tr>
<td>SGSN</td>
<td>Serving GPRS Support Node</td>
</tr>
<tr>
<td>SON</td>
<td>Self-Organizing Networks</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>TDD</td>
<td>Time Division Duplex</td>
</tr>
<tr>
<td>TFCI</td>
<td>Time Format Combination Indicator</td>
</tr>
<tr>
<td>TFI</td>
<td>Time Format Indicator</td>
</tr>
<tr>
<td>TTI</td>
<td>Transmission Time Interval</td>
</tr>
<tr>
<td>UARFCNDL</td>
<td>UTRA Absolute Radio Frequency Channel Number for the Downlink</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunication Services</td>
</tr>
<tr>
<td>UL</td>
<td>Uplink</td>
</tr>
<tr>
<td>USIM</td>
<td>UMTS Subscriber Identity Module</td>
</tr>
<tr>
<td>UTRAN</td>
<td>UMTS Terrestrial Radio Access Network</td>
</tr>
<tr>
<td>UTRA</td>
<td>Universal Terrestrial Radio Access</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>VCC</td>
<td>Virtual Channel Connections</td>
</tr>
<tr>
<td>VLR</td>
<td>Visitor Location Register</td>
</tr>
<tr>
<td>WCDMA</td>
<td>Wideband Code Division Multiple Access</td>
</tr>
<tr>
<td>WGS84</td>
<td>World Geodetic System</td>
</tr>
</tbody>
</table>
# Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>traffic density</td>
</tr>
<tr>
<td>$A_{CC}$</td>
<td>accessibility</td>
</tr>
<tr>
<td>$b$</td>
<td>resource consumption</td>
</tr>
<tr>
<td>$C$</td>
<td>available channels</td>
</tr>
<tr>
<td>$C_{CCh}$</td>
<td>common channel code usage</td>
</tr>
<tr>
<td>$c_a$</td>
<td>connection attempts</td>
</tr>
<tr>
<td>$D_{av}$</td>
<td>average delay</td>
</tr>
<tr>
<td>$I_{capacity}$</td>
<td>capacity imbalance factor</td>
</tr>
<tr>
<td>$I_{result}$</td>
<td>resulting imbalance factor</td>
</tr>
<tr>
<td>$I_{usage}$</td>
<td>usage imbalance factor</td>
</tr>
<tr>
<td>$L$</td>
<td>cell load</td>
</tr>
<tr>
<td>$L_{avg}$</td>
<td>average neighbour cell load</td>
</tr>
<tr>
<td>$L_{diff}$</td>
<td>load difference</td>
</tr>
<tr>
<td>$L_v$</td>
<td>quantizer level</td>
</tr>
<tr>
<td>$M$</td>
<td>number of subscribers</td>
</tr>
<tr>
<td>$m$</td>
<td>modulation factor</td>
</tr>
<tr>
<td>$N_{events}$</td>
<td>number of events</td>
</tr>
<tr>
<td>$N_M$</td>
<td>Markov chain state</td>
</tr>
<tr>
<td>$N_{result}$</td>
<td>resulting number of events</td>
</tr>
<tr>
<td>$N_{target}$</td>
<td>target number of events</td>
</tr>
<tr>
<td>$N_u$</td>
<td>number of users</td>
</tr>
<tr>
<td>$n_{CE}$</td>
<td>channel element usage</td>
</tr>
<tr>
<td>$O_{SH}$</td>
<td>soft handover overhead</td>
</tr>
</tbody>
</table>
$O_D$  
DPCCH overhead

$P$  
probability

$P_B$  
blocking probability

$P_N$  
$N$ active calls probability

$P_S$  
$N$ simultaneous connections probability

$p$  
coding puncturing

$R_b$  
bit rate

$R_c$  
chip rate

$R_{CCh}$  
channel coding rate

$S_F$  
spreading factor

$T$  
handover threshold

$T_{base}$  
base threshold

$T_{reg}$  
regulative threshold

$T_{RSCP}$  
RSCP threshold

$t$  
time

$\alpha$  
activity factor

$\Gamma_{DL}$  
downlink channel element factor

$\Gamma_{UL}$  
uplink channel element factor

$\Delta I$  
imbalance differential

$\Delta I_{result}$  
resulting imbalance differential

$\lambda$  
ON to OFF rate

$\lambda_c$  
call request rate

$\lambda_N$  
Markov chain birth rate

$\mu$  
OFF to ON rate

$\mu_c$  
call success rate

$\mu_N$  
Markov chain death rate

$\rho$  
call arrival to outgoing ratio

$\rho_c$  
average call time
Chapter 1

Introduction

This chapter presents an overall view of this thesis and serves as an introduction, including the motivations that have driven its development, a summary of its structure and the various approaches taken for solving the multiple problems it presented.

1.1 Motivation

The Telecommunications industry is subjected to many expenses, coming either from day-to-day operation costs, usually called Operational Expenditure (OpEx), or from investments that need to be made in order to improve the network, Capital Expenditure (CapEx). Technological progress in the field, namely the appearance of more powerful user devices such as smartphones that allow higher bit rates, demands an increase in network capacity. Just last year, the total mobile data traffic grew 69% [1]. It is expected that by the end of 2021 data traffic will have increased ten-fold [2]. This causes the need for operators to keep investing in modernizing their networks and increasing the capacity of their sites, which means expenditure at the CapEx level.

In order to keep CapEx, and also OpEx, to a minimum, operators should make the most out of the available resources. Mobile operators need to be able to define network capacity in terms of useful costumer-centric Key Performance Indicators (KPIs), install capacity in the several network elements on a just-in-time basis and exploit soft capacity properties of modern network technologies [3]. Over and under dimensioning the network capacity are two situations
to be avoided. Thus, it becomes necessary to know how mobile traffic behaves and how it is evolving.

Densification of the networks in order to meet these traffic demands causes a greater level of complexity when optimizing the network parameters. The only way this optimization can be cost-efficient is to have more automated and autonomous systems such as Self-Organizing Networks (SON) [4]. By analysing the state of each site or frequency band, the software selects a set of optimal parameters, allowing a better distribution of the traffic depending on the available resources and their usage. This method of distributing traffic is defined as Load Balancing. Implementing a SON allows savings in expenditure for operators, for example by delaying the need to upgrade the capacity of a site that otherwise would be considered at its limit.

1.2 Approach

To present a solution for the problems mentioned above, it was decided to develop two algorithms: a Traffic Forecasting algorithm, which will allow the operator to know when to perform a capacity upgrade, adding radio equipment hardware in base stations in order to increase the traffic throughput of a site; and a Load Balancing algorithm that would allow this moment to be delayed as most as possible.

This thesis is developed considering the Third Generation (3G) Wireless Access Networks, which operates in two different frequency bands. The first one is the 2100 MHz frequency band, which was originally developed for the introduction of this technology, and the second is the 900 MHz band, which was adapted from the old Second Generation (2G) technology by reframing the initial bandwidth and attributing part of it to 3G. Operators usually don’t distribute traffic equally between these two bands, creating an overload of the 2100 MHz band. This band is usually given a higher priority for connections since the 900 MHz band is used mostly to assure coverage of the site. The Load Balancing algorithm developed will try to correct this issue.

In order to develop a meaningful, realistic and technically accurate solution, a collaboration with a Portuguese telecommunications consulting company, Celfinet, was established. Celfinet provided the input data necessary for the development of this thesis as well as guidance and supervision during the various stages of the project. This allowed designing a semi-automatic
algorithm for network optimization, thus contributing for the development of Self Organizing Networks in Portugal.

1.3 Structure

The thesis is divided in six chapters. The first presents a small introduction to the project, stating its motivation and goals. The second chapter presents a short summary of the basic concepts necessary to understand the developed work. The third will present a solution for the Forecast algorithm and its validation by calculating the associated prediction error. Chapter 4 presents a description of the Load Balancing algorithm developed, including the strategy used and its results. The fifth chapter presents a combination of both developed algorithms in order to verify how they can be used to delay reaching the capacity limit of sites. The last chapter contains the final remarks and conclusions as well as potential future improvements to be made.
Chapter 2

State of the Art

This chapter will present a summary of the main concepts and theoretical background needed to understand the work developed in this thesis. Firstly, an introduction about the 3G Wireless Access Networks is presented, including its architecture and functionality. Afterwards, an overview of the main work developed in Load Balancing and Traffic Analysis is detailed.

2.1 3G Wireless Access Networks

In this section the main 3G Wireless Access Network specifications, as detailed in [5], shall be defined, particularly the Universal Mobile Telecommunication Services (UMTS) which is more relevant for this thesis.

2G systems, like the Global System for Mobile Communications (GSM), were originally developed for efficient delivery of voice services. With the growing demand for delivery of other types of services, like video streaming and data transfer, it became necessary to design a new kind of Wireless Access Network. Initially, a solution was developed for the 2G systems, the General Packet Radio System (GPRS), but the growing bit rate demands made it clear that a deeper change was in order. Thus, it began the development of a 3G system.

The UMTS networks are one of the 3G systems and they allow a flexible delivery of many types of services without requiring a particular network optimization for each one. The
Wideband Code Division Multiple Access (WCDMA) solution used in UMTS brings the following capabilities:

- High bit rates;
- Low delays;
- Seamless mobility for packet data applications;
- Quality of Service (QoS) differentiation for high efficiency of service delivery;
- Simultaneous voice and data transfer;
- Interworking with the existing GSM/GPRS networks.

2.1.1 Radio Interface

The WCDMA technology has emerged as the most widely adopted third-generation radio interface. It was specified in the 3rd Generation Partnership Project (3GPP), which is a joint project between standardization bodies from Europe, Japan, Korea, the USA and China.

Main WCDMA Parameters

This section details the main characteristics of the WCDMA system, as presented in [5]. WCDMA is a wideband Direct-Sequence Code Division Multiple Access (DS-CDMA) system. This means that the user information bits are multiplied with a sequence of quasi-random bits (called chips) derived from CDMA spreading codes, in order to obtain a wide bandwidth spread signal. The use of a variable spreading factor and multi-code connections is supported to allow higher bit-rates.

The used chip rate is of 3.84 Mcps, for a bandwidth of approximately 5 MHz. This is considerably wider than most second-generation systems, that used a narrowband of about 200 kHz. Wider bandwidths allow not just higher bit-rates but also performance benefits, such as increased multipath diversity. Network operators can deploy several 5 MHz carriers to increase capacity.
WCDMA supports variable user data rates, thus allowing the concept of Bandwidth on Demand. The user data rate is constant during each time frame but it can be changed from frame to frame. The network controls this fast allocation of capacity in order to achieve optimal throughput for packet data services.

Two basic duplexing techniques are supported: Frequency Division Duplex (FDD) and Time Division Duplex (TDD). In FDD mode, two separate 5 MHz carriers are used for uplink and downlink, whereas in TDD only one carrier is time-shared between uplink and downlink.

The operation of asynchronous base stations is supported, unlike most second-generation systems, which eliminates the need for a global time reference, such as GPS. This way, the deployment of indoor micro base stations becomes easier.

WCDMA uses coherent detection based on the use of pilot symbols or common pilot on both uplink and downlink, as opposed to second-generation systems that do not use it for the uplink. This results in an increase of coverage and capacity on the uplink.

The air interface is designed in a way that allows advanced CDMA receiver concepts, such as multi-user detection and smart adaptive antennas, to be implemented by the network operator to obtain an increase in capacity and coverage.

WCDMA was developed to be deployed in conjunction with GSM, making it possible to execute handovers between both systems, thus taking advantage of the GSM coverage for the introduction of the new system.

**Spreading and Despreading**

As previously mentioned, user data with rate $R_b$ is multiplied by a spreading code. This process is called spreading. Each user data bit is multiplied by a sequence of $N$ chips, resulting in a spread signal of rate $N \times R_b$ with the same random, noise-like appearance of the spreading code. This wideband signal is then transmitted.

At reception, the signal goes through a process of despreading. This is done by multiplying the received signal by the same spreading code used at transmission. If we have perfect
synchronization between the received signal and the spreading code, the original signal can be perfectly recovered.

Figure 2.1 shows an example of both these operations, with \( N = 8 \).

![Figure 2.1: Spreading and Despreading in DS-CDMA](image)

The correlation receiver for CDMA integrates the resulting despread signal for each user data bit. If this operation is applied to a signal from a different user, with a different spreading code, the result of the integration leads to interfering signal values fluctuating around zero. This process is illustrated in Figure 2.2.

![Figure 2.2: DS-CDMA correlation receiver principle](image)

The amplitude of the original user’s signal is increased by a factor equal to the spreading factor relative to the signal of the interfering signal. This effect is called processing gain and
is fundamental to CDMA systems since it gives robustness against self-interference. This is a great advantage of this type of systems since it allows the re-utilisation of frequencies over short distances.

This type of correlation receivers are widely used in base stations and terminals. However, due to multipath propagation and multi antenna receivers, it becomes necessary to have multiple correlation receivers in order to recover the energy from all paths and antennas. This array of correlation receivers are usually called fingers and constitutes the CDMA Rake receiver.

A consequence of the spreading and despreading process is that the transmission bandwidth is increased by an amount equal to the processing gain. When examined at a system level, it can be concluded that, due to the wideband properties of the signals, the WCDMA allows the following benefits:

- Frequency reuse of 1 between different cells, thus increasing spectral efficiency;

- Interferer diversity, which means that, since many users share the same wideband carrier, the multiple access interference from many users is averaged out. This allows an increase in system capacity compared to systems planned for the worst-case interference;

- The different propagation paths can be resolved at higher accuracy than in a system with lower bandwidth signals, allowing higher diversity. This improves performance by reducing the fading effect.

However, these benefits require the use of tight power control and soft handover to avoid one user’s signal blocking the others’. The power control used in WCDMA causes mobiles to increase or decrease their transmission power according to the air interface conditions. A direct consequence of this power control is a higher average transmission power for mobiles [5].

Soft Handover allows a mobile to be connected to two different air interfaces simultaneously. This presents a gain through diversity, but requires additional resources from the system such as more rake receiver channels in base stations, more transmission links between base station and Radio Network Controller (RNC), and more rake fingers in mobiles [5].
2.1.2 Radio Access Network Architecture

The UMTS consists of several network elements, where each has its own functionality. These network elements are defined at the logical level, but often they have a similar physical implementation. This is especially due to the fact that some of the interfaces are open, which means that they are defined to such a detailed level that the equipment at both ends can be from different manufacturers.

The network elements can be grouped into the Radio Access Network (RAN), or UMTS Terrestrial RAN (UTRAN), which is responsible for all radio-related functionality, the Core Network (CN) that is in charge of switching and routing calls and data connections to external networks, and the User Equipment (UE) that connects the user to the network. This architecture is shown in Figure 2.3.

Another way of grouping the UMTS network elements is to divide them into sub-networks, for instance the UMTS Public Land Mobile Network (PLMN). Typically, one PLMN is operated by one operator and is connected to other PLMNs and to external networks. Figure 2.4 illustrates the detailed network architecture of UMTS (after Release 4) and in Figure 2.5 the UTRAN, UE and interface architecture is detailed.

The UE is divided in two elements: the Mobile Equipment (ME) that is the radio terminal used for communication over the $Uu$ interface; and the UMTS Subscriber Identity Module (USIM), which is a smartcard that holds the user’s identity, performs authentication algorithms and stores information needed by the terminal.

The UTRAN is also composed of two distinct elements: the Node B that switches the data flow between the $Uu$ and $Iub$ interfaces and is also responsible for some resource management functionality; and the Radio Network Controller (RNC) that controls the radio resources in its
influence, i.e. the Node Bs connected to it, and also connects the UTRAN to the CN through the Iu interface.

An UTRAN can consist of one or several Radio Network Sub-systems (RNS), where each RNS is a sub-network with one RNC and one or more Node Bs. This architecture is specified with more detail in Figure 2.5. The main characteristics of UTRAN, which are also the main requirements for the design of its architecture, functionality and protocols, are:

- Support of UTRA and related functionality, mainly the support of soft-handover and WCDMA Radio Resource Management algorithms;
- Maximization of the similarities in the handling of PS and CS data, namely using the same interface to communicate with both PS and CS domains of CN;

- Use of Asynchronous Transfer Mode (ATM) as the main transport mechanism;

- Use of the Internet Protocol (IP) based transport as an alternative mechanism.

As previously mentioned, the RNC is responsible for the control of radio resources of the UTRAN and is a connection point to the CN. It is also responsible for data encryption and for terminating the Radio Resource Control (RRC) protocol, which defines messages and procedures between the UE and the UTRAN.

The RNC controlling a Node B is designated as its Controlling RNC. It is responsible for the load and congestion control of its Node Bs and for admission control and code allocation for new UEs trying to connect to that Node B.

For UEs connecting to more than one RNS, the RNCs involved have two different logical roles:

- The Serving RNC terminates both the Iu link for the transport of user data and the corresponding RAN application part signalling to the CN. It also terminates the RRC signalling between the UE and UTRAN. It is also responsible for basic Radio Resource Management (RRM) operations such as mapping Radio Resource Access Bearer (RAB) parameters, handover decision and outer loop power control;

- The Drift RNC is any RNC, other than the Serving RNC, that controls Node Bs used by the UE. It may perform macro-diversity combining and splitting. It may also route the data between the Iub and Iur interfaces.

The Node B, also commonly referred to as Base Station, performs some signal processing, like channel coding and interleaving, rate adaptation, and spreading. It is also responsible for some basic Radio Resource Management operations such as inner loop power control.

The main elements of the CN are similar to the ones defined in the GSM CN:
• The Home Location Register (HLR) is a database that stores the master copy of the user’s service profile, in the user’s home system;

• The Mobile Services Switching Centre/Visitor Location Register (MSC/VLR) are the switch and database, respectively, that serve the UE in its current location for Circuit-Switched (CS) services;

• The Gateway MSC (GMSC) is the switch at the point where the CN is connected to external CS networks;

• The Serving GPRS Support Node (SGSN) is the Packet-Switched (PS) equivalent of the MSC/VLR;

• The Gateway GPRS Support Node (GGSN) is the PS equivalent of the GMSC.

• The Media Gateway (MGW) performs the actual switching of user data and network interworking processing, i.e. echo cancellation or speech decoding/encoding.

The interfaces between each network element are defined as:

• The $Cu$ interface is the electrical interface between the USIM and the ME, and it demands a standard format for the smartcard;

• The $Uu$ interface is the WCDMA radio interface through which the UE is connected to the UTRAN;

• The $Iu$ interface connects the UTRAN to the CN and it allows the network operators to acquire UTRAN and CN equipment from different manufacturers;

• The $Iur$ interface allows soft handover between RNCs from different operators;

• The $Iub$ interface connects Node Bs to an RNC.

2.1.3 Frequency Bands

Initially the 2100 MHz frequency band was attributed to UMTS, more specifically the $[1920, 1980]$ MHz band for the UL and the $[2110, 2170]$ MHz for the DL. Later on, the 3GPP has
defined a new frequency band in the 900 MHz GSM band, where each carrier would have a 5 MHz bandwidth with a 200 kHz guard band in each side [6]. However, this reframing caused the total GSM bandwidth, and thus capacity, to be reduced.

This new band allows better radio propagation, since it operates at a lower frequency than the original 2100 MHz UMTS band. As a direct consequence, this is a more efficient option to cover low density populated areas. Due to this reason, the new frequency band is used by operators to assure the coverage of sites causing it to be seldom overloaded, whereas the 2100 MHz band is usually on the limit of capacity. The coverage gains are presented in Table 2.1 [6].

**Table 2.1:** Comparison between UMTS 900 MHz and UMTS 2100 MHz bands [6].

<table>
<thead>
<tr>
<th>Environment</th>
<th>Dense Urban</th>
<th>Urban</th>
<th>Suburban</th>
<th>Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>900 MHz vs. 2100 MHz</td>
<td>87%</td>
<td>44%</td>
<td>60%</td>
<td>119%</td>
</tr>
</tbody>
</table>

2.1.4 Handover

In this section the handover process will be described, based on [5, 6]. The handover consists on transferring an ongoing call or a data session from one channel to another, provided that the new channel offers a better radio connection. It can be triggered by several reasons, like user mobility or load balancing. In UMTS there are three types of handover:

- **Hard Handover:** All previous radio links are removed before connecting to the new ones. It may be non-seamless or seamless, meaning that it might be perceptible to the user or not, respectively. Inter-frequency handover is an example of hard handover;

- **Soft Handover:** Radio links are added and removed in a way that the UE has always at least one active radio link. Intra-frequency handover to another cell is usually a soft handover;

- **Softer Handover:** A special case of soft handover where the new radio links belong to the same Node B.

During the soft handover, the UE is connected to several cells simultaneously, thus it becomes necessary to define the following sets:
• Active Set: Cells which signals are used during a soft handover. The cells eligible to be part of an active set are found by analysing its Received Signal Code Power (RSCP) or Common Pilot Channel (CPICH) $E_c/I_0$ (power to interference ratio);

• Monitored Set: Cells belonging in the UE cell info list but not yet added to the active set;

• Detected Set: All the other remaining discovered cells.

2.1.5 Transport Channels

The data generated by users is carried through transport channels, which are mapped in the physical layer to physical channels. The physical layer must be able to support variable bit rate transport channels to offer bandwidth on demand services.

Each transport channel has an associated Transport Format Indicator (TFI) and the information of all TFIs is combined in the Transport Format Combination Indicator (TFCI). The TFCI is transmitted in the physical control channel to inform the receiver of which transport channels are being used in the current time frame. The TFCI is decoded at the receiver and the obtained TFI is given to the higher layers for each transport channel.

One physical control channel plus one or several physical data channels make up a Coded Composite Transport Channel (CCTrCh). Transport channels can be dedicated to one user or be shared between several users in a cell.

There is only one dedicated channel specified in UTRA, which is called the DCH. It transports all the data information for a certain user, as well as all the required control information such as handover commands and reports from the terminal. The DCH supports features such as fast power control, fast data rate change and soft handover.

There are six common channel types defined for UTRA in Release 99:

• The Broadcast Channel (BCH) is used to transmit information specific to the UTRA network or for a given cell, like the available random access codes and access slots in the cell. Since the terminal needs this information to register to the cell, this channel is
transmitted with relatively high power. The information rate in the BCH is limited by the ability of terminals to decode the date rate, resulting in a low and fixed rate;

- The Forward Access Channel (FACH) is a downlink transport channel that carries control information to the terminals known to be located in a cell. It is also possible to transmit packet data over this channel. There can be more than one FACH in each cell, but one of them as such a low bit rate that it might be received by all the terminals in the cell;

- The Paging Channel (PCH) is a downlink transport channel that carries data relevant to the initiation of communication between the network and the terminal. The terminals must be able to receive paging information in the whole cell. The design of the PCH affects the terminal’s power consumption in standby mode;

- The Random Access Channel (RACH) is an uplink transport channel used to carry control information such as requests to set up a connection. It can also be used for small packet data transport. Since it needs to be received inside the whole cell for proper system operation, the bit rate is rather low;

Depending on the type of traffic a user is generating, different transport channels are used for the transmission. For CS traffic the DCH is used, whereas for PS traffic, RACH is used for uplink and FACH for downlink.

In Release 5 the 3GPP introduced the High-Speed Downlink Packet Access (HSDPA) technology and defined a new common transport channel, the High-Speed Downlink Shared Channel (HS-DSCH) as well as two other control channels: High-Speed Shared Control Channel (HS-SCCH), which carries control information for HS-DSCH; and Uplink High-Speed Dedicated Physical Control Channel (HS-DPCCH), which carries control information for the uplink [5, 7]. The following changes were also implemented in Release 5:

- The Transmission Time Interval (TTI) has been reduced to 2 ms in order to achieve a short round-trip delay between the UE and Node B;

- A higher modulation scheme, 16QAM, and lower encoding redundancy were added to allow a higher peak data rate;
• The SF is fixed at 16, allowing the allocation of 15 codes to the users, since there is a need to have code space available for common channels, HS-SCCHs and the associated DCH.

In Release 6, the High-Speed Uplink Packet Access (HSUPA) technology is introduced. One of the major improvements made is the improvement of the DCH into the E-DCH, as well as the addition of the following channels [5, 7]:

• Enhanced DPDCCH (E-DPDCCH), that carries user data in the uplink direction;
• Enhanced DPCCCH (E-DPCCCH) which carries the E-DPDCCH related information, such as data rate information, retransmission information and scheduling control information;
• E-DCH Hybrid ARQ Indicator Channel (E-HICH), that carries packet reception success/-failure information in the downlink direction;
• E-DCH Absolute Grant Channel (E-AGCH) and E-DCH Relative Grant Channel (E-RGCH), that carry scheduling control information to control the uplink transmission rate.

In Release 7 the 3GPP introduced the High-Speed Packet Access Evolution (HSPA+), which introduced several improvements to end-user performance, network capacity and architecture [5]. One of the main improvements introduced was the multi-stream Multiple Input Multiple Output (MIMO) technology, which allows much higher bit rates. At the physical layer level the FACH, which remained unchanged in Releases 5 and 6, is replaced by an upgraded version of itself, the Enhanced FACH (eFACH).

The different transport channels are mapped to different physical channels, although some of the transport channels can be carried by the same physical channel. Some of the physical channels only carry information important to physical layer procedures and thus do not carry any transport channels. The DCH is mapped onto two physical channels, one of them carries the data and the other the necessary control information. This mapping is detailed in Figure 2.6
2.1.6 Downlink Cell Capacity

In WCDMA the downlink air interface capacity might be lower than the uplink capacity, depending on the load circumstances. This is mostly due to the fact that better receiver techniques can be used in the Node B than in the UE. Moreover, the downlink capacity is an important limitation because of the asymmetric nature of data traffic. The two main factors that limit the downlink capacity are the number of orthogonal codes available and the performance gain of the downlink transmit diversity.

The number of available downlink orthogonal codes within one scrambling code are limited by the spreading factor. With a spreading factor of $SF$, the total number of orthogonal codes available is also $SF$. This causes an upper limit on the downlink capacity. To obtain the maximum data throughput, besides the SF, the following factors must be known:

- Soft handover overhead;
- Chip rate (usually 3.84 Mcps);
- Modulation used;
- Common channels overhead;
• DPCCH overhead;

• Channel coding rate.

In [5] an estimate of the obtainable downlink data throughput with one set of scrambling codes and for one sector is calculated. The assumptions taken for this estimate are presented in Table 2.2. Using Equation 2.1, where $R_b$ is the obtained bit rate throughput, $R_c$ is the chip rate, $S_F$ is the spreading factor (and also the total number of codes available), $C_{CCh}$ is the number of codes used by common channels, $O_{SH}$ is the soft handover overhead, $m$ is the modulation factor (bits per symbol), $O_D$ is the DPCCH overhead, $R_{CCh}$ is the channel coding rate and $p$ the puncturing used, the obtained throughput per sector is 2.5 Mbps.

**Table 2.2: Assumptions in the calculation of downlink throughput.**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spreading Factor</td>
<td>128</td>
</tr>
<tr>
<td>Soft handover overhead</td>
<td>20%</td>
</tr>
<tr>
<td>Chip rate</td>
<td>3.84 Mcps</td>
</tr>
<tr>
<td>Modulation</td>
<td>QPSK (2 bits per symbol)</td>
</tr>
<tr>
<td>Common channel codes</td>
<td>10</td>
</tr>
<tr>
<td>Average DPCCH overhead</td>
<td>10%</td>
</tr>
<tr>
<td>Channel coding rate</td>
<td>1/3 with 30% puncturing</td>
</tr>
</tbody>
</table>

$$R_b = R_c \times \frac{S_F - C_{CCh}}{S_F} \times \frac{1}{1 + O_{SH}} \times m \times O_D \times R_{CCh} \times \frac{1}{1 - p}$$

Concerning the second factor mentioned earlier, the WCDMA standard allows the use of antenna diversity in the Node B, using for example two transmission antennas, in order to improve the downlink capacity. The performance gain obtained through this transmit diversity can be divided in two parts: coherent combining gain and diversity gain against fast fading. Coherent combining gain is obtained because the user signal from the two antennas is combined coherently, while interference is combined non-coherently. Gain against fast fading is obtained due to the uncorrelated fading obtained from the two transmission antennas [5].
2.1.7 Capacity Management in UMTS

As in [8], CS and PS traffic demand resources at different RAN levels: radio interface (Spreading Codes - SC), baseband processing capacity (Channel Elements - CE) and Iub capacity. While the SC and CE resources are shared between CS and PS services, the Iub interface is differentiated in Iub_CS and Iub_PS.

In the case of a resource shortage, more resources need to be allocated in the system. This can be performed at the several levels discussed before:

- At radio interface level, enhancements can be made by adding new carriers. Each 5 MHz carrier contains a SC tree;
- At baseband processing level, or Node B level, enhancements are made by installing new hardware baseband cards that provide extra CEs. Note that one CE is the baseband processing capacity required at the Node B for one voice channel;
- At Iub level upgrading the capacity involves increasing the number of Virtual Channel Connections (VCCs) when using ATM (CS services), or increasing IP throughput when using IP (PS services).

It is essential that network operators are capable of correctly identifying the causes of service degradation in their systems. This is not always an easy task and for this purpose a number of traffic models have been developed.

Channel Elements

In this section, a more detailed explanation of CEs and their impact on system capacity will be given, based on [9]. As said before, the CE is a measure of the required resource allocation in the Node B necessary to provide capacity for one voice channel, including control plane and signalling.

The number of required CEs for a certain user depends on the traffic type of the services he is using. Each service has a pair of channel element factors that characterize it, one for the
uplink ($\Gamma_{UL}$) and another for the downlink ($\Gamma_{DL}$). The CE capacity in a Node B depends on the hardware baseband cards installed and on the software licensing restrictions and it is shared between the several sectors of the cell. However, if the site has more than one frequency layer, each one has its own independent CE pool. In [10] a list of several baseband hardware cards is presented along with the CE capacity available in each one.

Knowing the number of simultaneous users, $N_{ui}$, in a site using service $i$ and the channel element factors, $\Gamma_i$, of each service, the total number of required CEs in a cell, $n_{CE}$, can be obtained through Equation 2.2.

\[
n_{CE} = \begin{cases} 
  \sum_i N_{ui} \Gamma_{iUL}, & \text{for UL} \\
  \sum_i N_{ui} \Gamma_{iDL}, & \text{for DL}
\end{cases}
\]  

(2.2)

2.2 Load Balancing

This section will address the load balancing problematic in 3G wireless access networks, based on [4]. An interesting way of making networks more efficient is to treat both UMTS frequency layers as a single global resource, instead of considering them independent resource pools. This aspect makes possible defining new load balancing strategies taking into account the requirements of the several available services the operator provides.

It is shown in [11] that when the channel element resource utilization reaches about 70% of the total capacity, the access success rate of each service decreases rapidly. This value can be used as a threshold for switching users from a determined frequency band to another, enhancing the quality of the provided services.

2.2.1 Capacity Driven Load Balancing

Capacity problems can be solved by adding hardware to the existing installed equipment, but this is often not the most economically efficient way. Alternatively, most infrastructure manufacturers offer some Radio Resource Management (RRM) features like call resource reduction mechanisms.
(causing call quality degradation or blocking), blind offloading (facilitating or forcing outgoing HOs without considering the impact on the performance of the target cell) and call pre-emption [4]. These mechanisms are essential for short-term capacity problems, but not for mid or long-term traffic growth issues.

A possible solution is to use an autonomous external optimization tool with a global vision of the network. This tool would apply a set of optimization policies which follow the operator’s traffic management strategy and propose a different way of assigning users to the different frequency layers, as illustrated in Figure 2.7.

![Figure 2.7: Example of inter-frequency load balancing][4]

In general, the parameters that control load balancing mechanisms can be divided into two categories:

- Cell level parameters, which affect the cell under consideration;
- Adjacency level parameters, which affect a certain neighbour relation. This category can be further divided into the two following types:
  - Intra-frequency neighbour related (IAF), for neighbours in the same frequency layer;
  - Inter-frequency neighbour related (IEF), for neighbours in different frequency layers.
UMTS Indicators Driving Load Balancing Processes

The utilization of traffic balancing policies between two layers or sectors is triggered when a significant imbalance is detected between them in terms of congestion or resource allocation. To detect the appearance of an imbalance, the optimizing tool must analyse some Key Performance Indicators (KPI) given by the network. The most relevant KPIs for the load balancing process are the following:

- DL transmitted power;
- Channelization code utilization;
- Blocking statistics;
- Congestion and rejection of bit upgrade request statistics;
- Iub bandwidth utilization;
- Channel element utilization;
- Call processor load at RNC.

UMTS Load Balancing Mechanisms

There are two main load balancing mechanisms that can be used in parallel:

- Cell Coverage Adjustment;
- Handover and Cell Reselection Adjustment.

The first can be done by remote electrical tilts and/or controlling the Common Pilot Channel (CPICH) power. In general this mechanism reduces the coverage area of highly loaded cells, automatically offloading traffic to less loaded neighbours. On the other hand, the coverage of cells with low load can be increased if their neighbours are highly loaded. However, this mechanism can introduce impacts on coverage and quality, so these must be carefully considered when designing the rules for the load balancing mechanism.
The second mechanism tunes handover and cell reselection thresholds, artificially shrinking or expanding cells by triggering handovers closer or farther from the initial point. As opposed to the previous mechanism, this one does not have the same implications on coverage or quality since pilot coverage is not impacted.

Neighbour specific handover parameters are usually defined as an offset to be added to the radio measurement of the neighbour cell that is evaluated for the handover decision. In this way, a neighbour cell can be either favoured or not in the handover decision by adding a positive or negative offset. The cell edge is thus redefined, allowing a full flexibility to adjust cell size and shape depending on the traffic distribution in the area. It is also important to verify that this method is less efficient and more limited to IAF neighbours, since border shifts can impact call quality due to interference.

IAF and IEF can be treated differently by using different threshold ranges, where a lower range imply fewer additional handovers and/or cell re-selections. Moreover, some neighbours can be favoured by configuring more appealing handover conditions in order to transfer more traffic towards them.

2.2.2 Dynamic Handover Algorithms

In [12] two algorithms are presented to provide a better solution to the original UMTS Fixed Threshold (FT) soft handover algorithm. The first one is the Distributed Adaptive Threshold (DAT) algorithm and the second the Concentrated Adaptive Threshold (CAT) algorithm.

In the DAT algorithm, each cell periodically updates its handover thresholds based on its load. When the load of a given cell is high, a lower threshold is set, and vice versa. Since different cells have different load levels, these should be separated into a finite number \( n \) of levels and each level \( L_{v_i} \) should be assigned to a threshold value \( T_i \) according to the quantizer \( Q \) given in Equation 2.3.

\[
T_i = Q(L_{v_i}), \quad i = 1, 2, \ldots, n
\] (2.3)
In the CAT algorithm, the threshold of each cell is set by the controlling RNC, based on the load of the cell itself but also in their neighbours’ load. To calculate the threshold of a given cell, the algorithm takes the load of the cell \( (L) \) and the average load of its neighbours’ load \( (L_{avg}) \) and calculates the load difference \( (L_{diff} = L - L_{avg}) \). Then it levels \( L \) and \( L_{diff} \) to \( L_{vi} \) and \( L_{vj} \), respectively, as in the DAT algorithm. Finally, both \( L_{vi} \) and \( L_{vj} \) are quantized to the base threshold \( (T_{base}) \) and regulative threshold \( (T_{reg}) \), respectively, and the final handover threshold \( (T) \) is calculated through Equation 2.4. Note that in this algorithm both quantizers, \( Q_1 \) and \( Q_2 \), may be different.

\[
T = T_{base} + T_{reg} = Q_1(L_{vi}) + Q_2(L_{vj})
\]  \hspace{1cm} (2.4)

The update period of adaptive thresholds are selected taking into consideration a compromise between the accuracy of the threshold setting and the threshold overhead. A short update period provides a more accurate setting of the thresholds, but a longer one provides less signalling overhead. In [12] the chosen update period was 100 ms.

### 2.3 Traffic Analysis

Data traffic is growing dramatically, especially since LTE was introduced. The high growth rates demand that mobile service providers anticipate and keep ahead of growing data traffic in order to make sure that their networks are able to handle this rapid growth [13]. For this purpose, several traffic models and forecasting algorithms were created, for instance the Mobile Data Forecasting and the Traffic Forecast Model created by Bell Labs and described in [13]. Some of these methods will be approached in this section.

#### 2.3.1 Erlang Traffic Models

The Erlang models are defined in [5, 14, 15]. It is a model for traffic density generated by voice calls. It can be obtained by using Equation 2.5, where \( A \) represents the traffic density, \( \lambda_c \) the call request rate and \( \mu_c \) the call success rate. Another way to calculate \( A \) is to consider the number of users \( N_u \) and the average call time \( \rho_c \), as explained in Equation 2.6.
\[ A[\text{Erlang}] = \frac{\lambda_c [\text{calls/s}]}{\mu_c [\text{calls/s}]} \]  
\[ A = N_u \frac{\rho_c [\text{s}]}{3600} \]

There are two main Erlang models, B and C. In the Erlang B model there is no queueing for call requests. If no channels are available the call is blocked. Knowing \( A \), the blocking probability \( P_B \) can be calculated using Equation 2.7, where \( C \) is the total number of available channels.

\[ P_B = \frac{A^C}{C!} \sum_{k=0}^{C-1} \frac{A^k}{k!} \]  

In the Erlang C model, if there are no channels available for a call request, then the call is queued for a certain amount of time. In this model, it is convenient to define the probability of a call being delayed, given by Equation 2.8. Knowing this probability, it is possible to calculate both the probability of a call being delayed more than \( t \) seconds, as given by Equation 2.9, and the average delay for all calls in the system, given by Equation 2.10.

\[ P(\text{delay} > 0) = \frac{A^C}{A^C + C! \left(1 - A/C\right) \sum_{k=0}^{C-1} \frac{A^k}{k!}} \]  
\[ P(\text{delay} > t) = P(\text{delay} > 0)e^{-C/A} \]  
\[ D_{av} = P(\text{delay} > 0) \rho_c \frac{C - A}{C} \]

In [8] a blocking probability model is detailed. The objective of modelling the blocking probability is to find out the network’s availability and knowing how to improve it. In UTRAN, a connection attempt is blocked when the necessary resources are not available. These resources
need to simultaneously provide different services. A more detailed view on this model is presented in Appendix A.

2.3.2 ON/OFF Model

Another common way to represent traffic is through an ON/OFF model [16]. When a source is in the ON state it generates packets with a constant inter arrival time, when it is OFF it does not generate any packets.

The voice source can be modelled as a two-state Markov chain, where the voice call process transits between ON and OFF states. The state transition diagram is shown in Figure 2.8 where $\lambda$ is the OFF to ON rate and $\mu$ is the ON to OFF rate. The average length of the ON and OFF periods are $1/\mu$ and $1/\lambda$ respectively.

![Figure 2.8: ON/OFF model [16].](image)

In this model, calls were generated according to a Poisson process with mean call duration of 120 seconds and activity and silence periods are generated by an exponential distribution with mean value of 3 seconds and independent on UL and DL. However, these parameters may be changed according to the user’s needs.

Considering an area with $M$ subscribers generating calls, users may be in the cell under study or outside of it and thus “arriving” to the cell. The time it takes for a user to arrive to the cell is given by a random variable with exponential distribution whose mean is $1/\lambda$. The call duration of users in the cell is also exponentially distributed with an average value of $1/\mu$. A continuous time birth-death Markov chain can be built, where its states are given by the number of users having a call in progress.
When $N_u$ users have a call in progress in the cell, the Markov chain is in state $N_M$ and the birth and death rates are $\lambda_N$ and $\mu_N$ respectively. These two rates are given by Equations 2.11 and 2.12.

$$\mu_{N_M} = \mu N_M, \quad N_M = 1, 2, ... \quad (2.11)$$

$$\lambda_N = \begin{cases} (M - N_M)\lambda, & 0 \leq N_M < M \\ 0, & \text{otherwise} \end{cases} \quad (2.12)$$

Under equilibrium, the probability of having $N_u$ users with an active call is given by Equation 2.13 where $\rho = \lambda/\mu$ is the arrival/outgoing ratio.

$$P_N = \frac{M}{N_u} \frac{\rho^{N_u}}{(1 + \rho)^{N_u}} \quad (2.13)$$

Finally, the probability of $N_u$ users transmitting simultaneously is given by Equation 2.14 where $M$ is the number of users camping on the cell, $N_u$ is the number of users transmitting in a given moment and $\alpha$ is the activity factor, probability of state ON, given in Equation 2.15.

$$P_S = \left( \frac{M}{N_u} \right) \frac{(\alpha \rho)^{N_u}(1 + (1 - \alpha)\rho)^{M-N_u}}{(1 + \rho)^M} \quad (2.14)$$

$$\alpha = \frac{E[ON_{duration}]}{E[ON_{duration}] + E[OFF_{duration}]} \quad (2.15)$$

### 2.3.3 Traffic Prediction Models

Traffic evolution is highly unpredictable since it depends on many different factors. In order to correctly plan and optimize their network, operators need to estimate traffic growth. For that motive, some traffic prediction models were developed.
Daily Traffic Prediction

The daily variation of mobile traffic can be modelled in two different ways, using a double Gaussian model or a Trapezoidal model, as described in [6]. Both of these models present a fitting method for forecasting.

The double Gaussian model considers that the traffic behaves in such a way that two peaks are created throughout the day, one in the morning and another in the afternoon. The double Gaussian function is given in Equation 2.16 where \( p_1 \) is the peak traffic at \( h_1 \), which is the morning peak hour, \( d_1 \) is the first Gaussian deviation. \( p_2 \) is the peak traffic at \( h_2 \), which is the afternoon peak hour, \( d_2 \) is the second Gaussian deviation. \( h_t \) is the transition point, where the two Gaussian functions coincide.

\[
 f_{Gauss}(t) = \begin{cases} 
 p_1 e^{-\frac{(t-h_1)^2}{2d_1^2}}, & t < h_t \\
 p_2 e^{-\frac{(t-h_2)^2}{2d_2^2}}, & t > h_t 
\end{cases} \tag{2.16}
\]

In the Trapezoidal model considers that there isn’t a defined peak hour but instead a peak period between morning and afternoon. The trapezoidal function is given in Equation 2.17 and it is constructed using two Gaussian functions like in Equation 2.16 but truncated between \( t_1 \) and \( t_2 \) where it has a maximum value of \( c \).

\[
 f_{Trapez}(t) = \begin{cases} 
 p_2 e^{-\frac{(t-h_1)^2}{2d_1^2}}, & t < t_1 \\
 c, & t_1 < t < t_2 \\
 p_2 e^{-\frac{(t-h_2)^2}{2d_2^2}}, & t > t_2 
\end{cases} \tag{2.17}
\]

Long Term Traffic Prediction

In [17], it is concluded that, unlike voice traffic, mobile data traffic exhibits statistical self-similarity and needs to be modelled using a Fractional Auto-Regressive Integrated Moving Average (FARIMA) process. However, this process is still not accurate enough for data traffic
prediction, since it may fail due to the multi-fractal characteristic of this type of traffic. To re-
solve this issue, the authors present a new method that combines FARIMA and Auto-Regressive
Moving Average (ARMA) to better predict data traffic in 3G mobile networks. The obtained
experimental results reveal the effectiveness of the new process.

Another proposed method is a seasonal Auto-Regressive Integrated Moving Average (ARIMA),
presented in [18], which gives accurate long-term forecasts. This method also provides predic-
tions with considerably low prediction error, as shown by the experimental results obtained.

In [19] entropy theory is used to analyse the feasibility of predicting traffic variation theo-
retically. This analysis validates the spatial and temporal relevancies in typical types of traffic.
Based on this analysis several methods for traffic prediction and its performances are presented.

Finding an optimal traffic prediction algorithm is a technical challenge, since there is
usually a trade-off between prediction accuracy, complexity and computational power. In [20],
several low complexity models are evaluated, mainly by comparing actual traffic values with
predicted ones, to find the better solution. Based on the results, a hybrid prediction model
that combines Linear Regression and Naive Forecasting is presented. This solution is based on
a fitting method, where several functions are tested to see which presents a better correlation
with the input data.
Chapter 3

Traffic Forecasting

This chapter presents a detailed characterization of the algorithm used for traffic prediction in this thesis. Firstly, the available input data parameters are detailed and analysed, including a weekly usage analysis of the sites. Then, the different stages of the algorithm are specified. Finally, the results are presented along with a detailed analysis.

3.1 Overview

As mentioned earlier, this chapter presents a solution for a traffic forecasting algorithm. The motivation for developing this algorithm was to help mobile operators realise when a capacity expansion is needed, by predicting the approximate evolution of traffic over time.

There are many different methods of constructing a forecast algorithm, several were presented in Section 2.3.3 but the chosen approach was based on a fitting method. This type of method allows a relatively short computation time due to its low complexity. It fits several pre-determined functions to the data set choosing the best one for forecasting. In the following sections this method is extensively explained.
3.2 Input Data

The input data used as the input for the forecast algorithm consists of network statistics and topology of a Portuguese Mobile Telecommunications operator, and it includes information from several sites in Lisbon. The available data consists of daily samples of voice and data traffic volume, average number of users and some QoS metrics as well as the characterization of each site. The most relevant indicator available is the CE usage at the Node B and it is presented in both average usage and maximum usage, which is defined as the usage in the busy hour of the corresponding day. The busy hour is considered to be the hour which presents the highest traffic volume and is usually used for dimensioning the network. A summary of the available data is presented in Table 3.1.

Table 3.1: Characteristics of available input data for traffic forecasting.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sites</td>
<td>43</td>
</tr>
<tr>
<td>Base Stations</td>
<td>86</td>
</tr>
<tr>
<td>Cells</td>
<td>407</td>
</tr>
<tr>
<td>Sample Frequency</td>
<td>Daily</td>
</tr>
<tr>
<td>Total Sample Period</td>
<td>192 days</td>
</tr>
</tbody>
</table>

In the aforementioned table, a Site is defined as a geographical location, where, in this case, two Base Stations are positioned, one for each frequency band, and cover the Site’s area. A Base Station is another designation for the Node B. A Cell is defined as a portion of the area served by each Base Station, and each Base Station may have one or several Cells. Each Cell only serves one carrier, so for Sites with several carriers there may be Cells overlapping in the same geographical area.

In order to obtain more significant results, since the data is highly periodical over each week, it was decided to also apply the algorithm to three different compressed data sets. These compressed sets make use of the data periodicity by grouping each 7 consecutive samples and creating a new array with a weekly sample frequency. The three methods used for compressing the data are defined in Equations 3.1 to 3.3:

- **Weekly Average:**

  \[ W_{avg} = \frac{1}{7} \sum_{k=t}^{7t+6} D_k \]  

  (3.1)
• Weekly Peak Value:

\[ W_{t_{pk}} = \max_{k \in [t, t+6]} (D_k) \]  

(3.2)

• Weekly 90\textsuperscript{th} Percentile:

\[ W_{t_{90}} = D'_i, \quad i = \left\lceil N \frac{P}{100} \right\rceil \]  

(3.3)

In the equations above, \( W_t \) is the weekly value obtained, where \( t \) indicates the corresponding week number (beginning in 0) and \( D \) is the original daily vector, where \( k \) represents the corresponding day number. In Equation 3.3, \( D' \) is a sorted vector of daily values for each week \( t \), from least to greatest value, \( N \) is its size, which is 7 since it is a weekly vector with values for each day, and \( P \) is the order of the percentile desired, which in this case is 90. In practice, since \( i \) equals 6, the 90\textsuperscript{th} Percentile will choose the second biggest value of the week.

3.2.1 RAN Topology File

In order to identify and locate the sites geographically and know its operating frequency band, it is necessary to have information concerning the RAN topology. This file contains all information regarding each cell element, including its location, ID and operating frequency. The most relevant parameters for this thesis and its descriptions are the following:

• **LocationId** - Site ID number, provides information regarding the co-location Node Bs, i.e. two Node Bs with the same site ID number are geographically co-located.

• **NodeId** - Node B ID string, provides information regarding the co-location of cells, i.e. two cells with the same Node B ID belong to the same Node B.

• **CELL\_REF** - Cell ID number, which identifies each cell with a unique designation.

• **SITE\_NAME** - Name of the geographical location of the Node B, plus frequency band name if operating frequency is in U900 band.

• **CELL\_NW** - Cell name, resulting of the corresponding **NodeId** plus a letter designating the cell.
• **UARFCNDL** - The UTRA Absolute Radio Frequency Channel Number for the Downlink is used to identify the operating frequency assigned to the operator in the cell. This is a code which translates to an uplink and downlink frequency and this correspondence can be made using the tool in [21].

• **LATITUDE** and **LONGITUDE** - Set of coordinates which characterize the geographical localization of the site, according to the World Geodetic System (WGS84) system.

### 3.2.2 Traffic Statistics Files

These files contain daily information regarding each cell or Node B. One of the files contains the CE average and maximum usage at a Node B as well as its total CE capacity. The other file contains a set of KPI measurements made at the cell level.

The available parameters at the Node B level are:

- **Datetime** - Time-stamp of each measurement.
- **NodeB** - Identifies the Node B where the measurement is made. Corresponds to the **NodeId** parameter of the RAN Topology file.
- **Basebandpool** - Indicates the baseband resource pool of the measurement, since Node Bs can have one or more baseband pools to increase capacity.
- **Avg\_Usage\_CEs\_DL** and **Avg\_Usage\_CEs\_UL** - Average CE usage during each sample period, i.e. 1 day, for the downlink and uplink respectively.
- **Max\_Usage\_CEs\_DL** and **Max\_Usage\_CEs\_UL** - Peak CE usage during each sample period for the downlink and uplink respectively.
- **Lic\_DL\_CE** and **Lic\_UL\_CE** - Total software limited CE capacity of the corresponding baseband resource pool at the time of the measurement, for the downlink and uplink respectively.

The most relevant parameters available at the cell level are:
• **Datetime** - Time-stamp of each measurement.

• **UTRAN_CELL** - Identifies the cell where the measurement is made. Corresponds to the **CELL_REF** parameter of the RAN Topology file.

• **KPI_Traf_Speech** - Total speech traffic volume, in Erlang, recorded at the corresponding cell for the duration of the sample period.

• **KPI_R99_DL_Traf** and **KPI_R99_UL_Traf** - Total Release 99 PS traffic volume, in Megabyte, for the downlink and uplink respectively recorded at the corresponding cell for the duration of the sample period.

• **KPI_HSDPA_Traf_Vol** - Total HSDPA traffic volume, in Megabyte, recorded at the corresponding cell for the duration of the sample period.

• **KPI_EUL_Traf_Vol** - Total EUL traffic volume, in Megabyte, recorded at the corresponding cell for the duration of the sample period.

• **Avg_Speech_Ue**, **Avg_PSR99_Ue** and **Avg_HSPA_Ue** - Average number of users per 15 minute Report Output Period (ROP) for each service, i.e. Speech, R99 PS and HSPA (HSDPA and EUL).

Other KPI measurements at this level include the Call Setup Success Rate (CSSR) and the Call Drop Rate (CDR) for each service and the Received Total Wideband Power (RTWP). In Table 3.2 a summary of the available input data is presented.

<table>
<thead>
<tr>
<th>Table 3.2: Summary of available input data for traffic forecasting.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metric</strong></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td><strong>Node B Level</strong></td>
</tr>
<tr>
<td>Average Usage</td>
</tr>
<tr>
<td>Maximum Usage</td>
</tr>
<tr>
<td>Capacity Limit</td>
</tr>
<tr>
<td><strong>Cell Level</strong></td>
</tr>
<tr>
<td>CSSR</td>
</tr>
<tr>
<td>CDR</td>
</tr>
<tr>
<td>Traffic Volume (Data)</td>
</tr>
<tr>
<td>Traffic Volume (Speech)</td>
</tr>
<tr>
<td>Average Users</td>
</tr>
<tr>
<td>RTWP</td>
</tr>
</tbody>
</table>
3.2.3 Data Parsing

The available data is stored in two types of files, Comma Separated Values (CSV) and Excel Worksheet Files, containing information which spans for several thousands of lines. These huge files are very demanding to work with on a computation level. In order of decreasing the complexity of the program and its runtime, the files were loaded into MATLAB®, converted to Table type variables and finally saved as MAT files. Afterwards, this allowed loading the MAT files only to get the input data, which consumes considerably less time than loading the CSV files.

3.3 Weekly Site Characterization

In order to understand how each site’s usage is distributed along each week, a function was developed to differentiate sites in three separate categories: working-days sites, weekend sites and sites with no weekly trend. This was achieved by comparing the average traffic, over the 192 day sample period, during weekends and week-days and considering a threshold of 10%, i.e. if the average usage in working-days is 10% greater than the average usage in weekends then the site is considered to be a working-days site.

Since most of the sites are located in the city centre the results are not very significant, however this function can be used to characterize a sample of sites with a larger disparity in weekly usage. The obtained results for the available sample of sites is detailed in Table 3.3. In Figure 3.1 the sites are portrayed in their actual geographic locations using Google Earth™. The sites marked in red are working-day sites and the ones marked in yellow present no weekly usage trend.

It can be observed that the sites marked in yellow are essentially located in touristic locations. Another relevant case to be mentioned is a site which covers one of the football stadiums. These sites do not have a defined weekly trend since the aforementioned situations are independent of the week day. The obvious exception is the football stadium, which has a high usage during certain weekends, usually every two weeks, and also on some working days when there are European competitions.
Figure 3.1: Mapping of sites according to their weekly usage.

Table 3.3: Site characterization according to weekly usage.

<table>
<thead>
<tr>
<th>Site Type</th>
<th>Number of Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working-days Site</td>
<td>35</td>
</tr>
<tr>
<td>Weekend Site</td>
<td>0</td>
</tr>
<tr>
<td>No Trend</td>
<td>8</td>
</tr>
</tbody>
</table>

These results can be used in forecasting traffic by selecting the more relevant set of data to predict. For example, if a site is considered a weekend site, then the forecasting algorithm can be adapted to predict only the weekend usage of the site, which will be the capacity limiting period. This can allow better prediction errors as well as more significant results for the operator, who needs to decide when and how to expand the capacity of the sites.

Unfortunately, due to the low diversity of sites available, these results are not relevant for the forecasting algorithm in this thesis. However, given a larger set of sites it may produce relevant results allowing lower prediction errors, thus improving the algorithm’s efficiency.
3.4 Forecasting Algorithm

The forecasting algorithm was developed using MATLAB® R2015a and it takes full advantage of the program’s capabilities. In order to keep the complexity and runtime low, a simple but efficient approach was taken. The algorithm has three stages, fitting, decision and validation. During the first two stages it tries to find the best fit for the given input data, whether it is traffic statistics or reported CE usage, to a certain curve function. On the third stage, it extends the best fit function to a future time, thus obtaining a forecast of the data. An overview flowchart of the algorithm is presented in Figure 3.2.

![Flowchart of the forecasting algorithm.](image)

**Figure 3.2:** Overview flowchart of the forecasting algorithm.

As mentioned in Section 3.2, the input data set spans over the duration of 192 days, with daily samples. In order to validate the forecasting algorithm, the data will be split into three smaller sets with equal duration. The first two sets will be considered as the real input for the algorithm, while the last one will be used to calculate the prediction error. This splitting is illustrated, with the graphic representation of the CE usage of the 2100 MHz frequency band in Site 2, in Figure 3.3, where section 1 is used for the fitting stage, section 2 for the decision stage and section 3 for the validation stage.
3.4.1 Fitting and Decision Stages

As mentioned before, the algorithm tries to fit the input data to a set of functions and chooses the best fit to make a forecast. The set of candidate functions chosen should reflect typical traffic behaviour in mobile networks. Taking into account this consideration, five different functions were chosen to find the best fit, these are detailed in Equations 3.4 to 3.8.

- **Linear Function:**
  \[ y = ax + b \]  \hspace{1cm} (3.4)

- **Quadratic Function:**
  \[ y = ax^2 + bx + c \]  \hspace{1cm} (3.5)

- **Power Function:**
  \[ y = ax^b + c \]  \hspace{1cm} (3.6)

- **Gaussian Function:**
  \[ y = ae^{-(x-b)^2 \over c^2}} \]  \hspace{1cm} (3.7)
Logarithmic Function:

\[ y = a + b \log(x) \]  (3.8)

In the previous equations, \( x \) represents the time variable, in either days or weeks, \( a \), \( b \) and \( c \) are the function parameters to be obtained in the fitting process, and \( y \) will be the resulting forecast obtained through the fitting process.

On these stages, the algorithm takes as input data the two first sub-sets mentioned in Section 3.4. Firstly, in the fitting stage, it fits the first sub-set to each of the five fitting functions using the MATLAB® fit or polyfit functions [22, 23]. The result is a fitobject variable which contains the set of calculated parameters \((a, b, c)\) of the fitting function.

Afterwards, in the decision stage, the algorithm makes a forecast spanning over the duration of the second sub-set of the input data using the MATLAB® predint or polyval functions [24, 25], that takes as an input the fitobject variable. Finally it compares the obtained forecast with the second sub-set of input data and finds the best fitting function for it, minimizing the prediction error. At this stage, the algorithm returns the best fitting function and its parameters for the validation stage. The flowchart of these two stages is presented in Figure 3.4.

### 3.4.2 Validation Stage

After obtaining the best fitting function for the data and its parameters it is possible to make a prediction using the predint or polyval functions, in a process similar to the decision stage of the algorithm. These functions calculate the fitting function’s values with the given parameters for a given future time span. Their output is a vector containing a prediction of the future value of the data for a certain number of given periods, where in this case each period is a single day or a whole week depending on the input data format.

In this thesis the considered future time span was the duration of the third sub-set of input data mentioned in Section 3.4. By comparing the actual values of the data with its prediction it was possible to obtain a quantitative measure of the algorithm’s accuracy. This comparison was made using a Mean Absolute Percentage Error (MAPE) and a Normalized Root Mean Squared...
Error (NRMSE), as explained in Equations 3.9 and 3.10, where $n$ is the total number of samples, $Y_i$ is the actual data value at time $i$ and $F_i$ is the forecast value at time $i$. In Equation 3.10, $\bar{Y}$ represents the mean value of vector $Y$.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - F_i}{Y_i} \right|$$

$$NRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \frac{(Y_i - F_i)^2}{\bar{Y}}}$$

The flowchart that illustrates this stage of the algorithm is presented in Figure 3.5.
3.4.3 Results and Analysis

The forecast algorithm was applied to several QoS metrics at the cell level as well as reported CE usage at the Node B level and an example of the obtained prediction can be seen in Figure 3.6. The input data used in the example was a vector of daily maximum UL CE usage in Site 2. In Figure 3.7 is illustrated an example for the same site but now with a weekly input data using the 90th percentile. In Figures 3.8 to 3.11 are examples of the forecast algorithm’s output when using the remaining metrics as input data.

Figure 3.6: Example of a forecast used for validation.
Figure 3.7: Example of a forecast used for validation, with weekly input data.

Figure 3.8: Example of a forecast used for validation, with speech traffic as input data.
Figure 3.9: Example of a forecast used for validation, with HSPA traffic as input data.

Figure 3.10: Example of a forecast used for validation, with R99 data traffic as input data.
In Tables 3.4 and 3.5 an overview of the obtained prediction errors, averaged over all sites, for the several cases of input data is presented, using MAPE and NRMSE, respectively.

Table 3.4: Average MAPE for the several input data types.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>CE Usage</th>
<th>Traffic Volume</th>
<th>Number of Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily Data [%]</td>
<td>Weekly Average [%]</td>
<td>Weekly Peak [%]</td>
</tr>
<tr>
<td>Maximum UL</td>
<td>24.46</td>
<td>13.52</td>
<td>11.69</td>
</tr>
<tr>
<td>Maximum DL</td>
<td>42.66</td>
<td>12.70</td>
<td>15.88</td>
</tr>
<tr>
<td>Average UL</td>
<td>32.80</td>
<td>18.93</td>
<td>16.88</td>
</tr>
<tr>
<td>Average DL</td>
<td>47.04</td>
<td>12.27</td>
<td>12.43</td>
</tr>
<tr>
<td>Speech</td>
<td>58.98</td>
<td>18.11</td>
<td>15.29</td>
</tr>
<tr>
<td>HSDPA</td>
<td>57.22</td>
<td>39.48</td>
<td>43.72</td>
</tr>
<tr>
<td>EUL</td>
<td>71.37</td>
<td>39.72</td>
<td>57.65</td>
</tr>
<tr>
<td>R99 DL</td>
<td>48.29</td>
<td>34.58</td>
<td>43.25</td>
</tr>
<tr>
<td>R99 UL</td>
<td>98.70</td>
<td>54.90</td>
<td>90.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metrics</th>
<th>CE Usage</th>
<th>Traffic Volume</th>
<th>Number of Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily Data [%]</td>
<td>Weekly Average [%]</td>
<td>Weekly Peak [%]</td>
</tr>
<tr>
<td>Speech</td>
<td>49.86</td>
<td>18.22</td>
<td>16.79</td>
</tr>
<tr>
<td>HSPA</td>
<td>36.05</td>
<td>24.83</td>
<td>23.88</td>
</tr>
<tr>
<td>R99</td>
<td>69.62</td>
<td>24.71</td>
<td>25.93</td>
</tr>
</tbody>
</table>

From the obtained results it is easily concluded that the weekly compressed data sets provide much better prediction errors, meaning it is easier to predict the weekly behaviour of the sites than its daily behaviour. Moreover, for most cases of input data, the 90th Percentile is the better aggregation method, minimizing the prediction error. However, in the Maximum DL CE usage, EUL and R99 DL Traffic Volumes present a slightly different trend, not having the
Table 3.5: Average NRMSE for the several input data types.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Daily Data [%]</th>
<th>Weekly Average [%]</th>
<th>Weekly Peak [%]</th>
<th>Weekly 90th Percentile [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CE Usage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum UL</td>
<td>24.94</td>
<td>15.04</td>
<td>13.67</td>
<td>12.75</td>
</tr>
<tr>
<td>Maximum DL</td>
<td>35.59</td>
<td>14.58</td>
<td>19.63</td>
<td>18.48</td>
</tr>
<tr>
<td>Average UL</td>
<td>30.35</td>
<td>20.59</td>
<td>18.15</td>
<td>18.74</td>
</tr>
<tr>
<td>Average DL</td>
<td>36.88</td>
<td>13.88</td>
<td>14.17</td>
<td>13.40</td>
</tr>
<tr>
<td><strong>Traffic Volume</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech</td>
<td>38.77</td>
<td>19.66</td>
<td>17.09</td>
<td>17.56</td>
</tr>
<tr>
<td>HSDPA</td>
<td>50.19</td>
<td>41.13</td>
<td>45.62</td>
<td>43.72</td>
</tr>
<tr>
<td>EUL</td>
<td>65.51</td>
<td>41.96</td>
<td>61.92</td>
<td>61.01</td>
</tr>
<tr>
<td>R99 DL</td>
<td>45.62</td>
<td>37.40</td>
<td>46.78</td>
<td>43.59</td>
</tr>
<tr>
<td>R99 UL</td>
<td>80.64</td>
<td>52.86</td>
<td>79.85</td>
<td>74.91</td>
</tr>
<tr>
<td><strong>Number of Users</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech</td>
<td>38.16</td>
<td>19.62</td>
<td>18.48</td>
<td>18.08</td>
</tr>
<tr>
<td>HSPA</td>
<td>34.12</td>
<td>26.59</td>
<td>25.51</td>
<td>25.73</td>
</tr>
<tr>
<td>R99</td>
<td>45.87</td>
<td>25.77</td>
<td>27.56</td>
<td>25.64</td>
</tr>
</tbody>
</table>

90th Percentile as the best method for aggregation of the input data. This may be due to the fact that these metrics present low usage, such as R99 DL Traffic, or highly irregular behaviour resulting in big prediction errors, as can be seen in Figures 3.8 to 3.11. In fact for these three metrics the best aggregation method is a weekly average, since this is more stable for irregular data than the 90th Percentile.

It is also possible to verify that the CE usage data is the easiest to predict, mainly because it is the most stable indicator. After that, the most reliable indicator seems to be the number of users statistics. Where it comes to traffic volume, the algorithm finds more difficulties due to its highly irregular behaviour along the sampling window. However, the speech traffic and number of users statistics seem to be a good indicator, having a considerably lower prediction error.

In order to evaluate the chosen functions’ usage and performance, a statistical test was performed. In Table 3.6 the usage of each function is presented, grouped by data type as well as a global view, which is an average of all input data cases. In Table 3.7 the same analysis is presented for the prediction errors (MAPE) obtained for each function.

From the previous tables it is possible to conclude that the logarithmic function is the most common choice of the forecast algorithm, whereas the Quadratic is the least chosen function. In terms of prediction error, the Quadratic function presents the worst performance, opposed to the Gaussian function that presents the best performance. There is a certain consistency
Table 3.6: Usage of each function in the forecast algorithm.

<table>
<thead>
<tr>
<th>Function</th>
<th>Daily Data [%]</th>
<th>Weekly Average [%]</th>
<th>Weekly Peak [%]</th>
<th>Weekly 90th Percentile [%]</th>
<th>Global View [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>23.49</td>
<td>14.53</td>
<td>13.84</td>
<td>12.56</td>
<td>16.10</td>
</tr>
<tr>
<td>Quadratic</td>
<td>6.16</td>
<td>3.37</td>
<td>8.84</td>
<td>8.95</td>
<td>6.83</td>
</tr>
<tr>
<td>Gaussian</td>
<td>22.67</td>
<td>28.26</td>
<td>23.02</td>
<td>24.30</td>
<td>24.56</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>37.56</td>
<td>41.28</td>
<td>41.16</td>
<td>40.81</td>
<td>40.20</td>
</tr>
</tbody>
</table>

Table 3.7: Prediction error associated with each function chosen by the forecast algorithm.

<table>
<thead>
<tr>
<th>Function</th>
<th>Daily Data [%]</th>
<th>Weekly Average [%]</th>
<th>Weekly Peak [%]</th>
<th>Weekly 90th Percentile [%]</th>
<th>Global View [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>48.41%</td>
<td>24.00%</td>
<td>26.00%</td>
<td>25.88%</td>
<td>33.70%</td>
</tr>
<tr>
<td>Quadratic</td>
<td>85.63%</td>
<td>52.37%</td>
<td>64.31%</td>
<td>58.80%</td>
<td>65.84%</td>
</tr>
<tr>
<td>Power</td>
<td>53.70%</td>
<td>30.98%</td>
<td>30.27%</td>
<td>29.77%</td>
<td>35.13%</td>
</tr>
<tr>
<td>Gaussian</td>
<td>44.58%</td>
<td>22.09%</td>
<td>25.15%</td>
<td>22.31%</td>
<td>28.05%</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>62.66%</td>
<td>30.74%</td>
<td>37.45%</td>
<td>36.66%</td>
<td>41.42%</td>
</tr>
</tbody>
</table>

between usage and performance, since the worst performing function is the least chosen and the best performing is the second most chosen function.

Note that the most chosen function does not always provide the lower prediction error. This is due to the fact that the fitting function is chosen in a way that minimizes the error in section 2 of Figure 3.3 and thus does not assure that it is the best choice for the third section.
Chapter 4

Load Balancing

This chapter presents a Load Balancing technique developed in order to better distribute the capacity usage between the two frequency layers, defining a new admission threshold for each site. This chapter presents the detailed process used for Load Balancing as well as each development decision and assumptions made.

4.1 Overview

As detailed in Section 2.2.1, there are several factors that may trigger a load balancing mechanism. In this thesis the main driving indicator will be the reported UL CE usage at each site. The presented algorithm analyses the CE usage in each layer, as well as the total allocated capacity, and based on this analysis, decides which sites may be improved.

Firstly, an imbalance factor is calculated for each site and, analysing it, the sites are selected for the Load Balancing process. Afterwards, some unavailable parameters need to be estimated due to lack of input data in the 900 MHz band. Finally, a suggested threshold is given as the final output of the algorithm. This algorithm was inspired on the work developed in [12], where the concept of dynamic thresholds is introduced. The algorithm’s efficiency is determined by recalculating the imbalance factors, considering that the network was optimized using the new thresholds.
4.2 Input Data

The data used as input for the Load Balancing algorithm contain traces of users from several sites in Lisbon, corresponding to some but not all sites used in Chapter 3 and only for one carrier in the 2100 MHz frequency band. These traces were positioned through an algorithm developed by Celfinet [26, 27]. The available information includes the geographic location of each measurement, the cell ID number, a time-stamp and several measurements such as the Received Signal Code Power (RSCP). The measurements were made on April 27th, 2015 between 10:00 and 11:00 AM.

| Table 4.1: Characteristics of available input data for load balancing. |
|-----------------|----------------|
| Sites           | 40             |
| Base Stations   | 40             |
| Cells           | 109            |
| Sample Frequency| Every second   |
| Total Sample Period | 1 hour       |

The complete detailed set of available measurements is the following:

- **UE_CONTEXT** - Unique ID number which identifies each UE.
- **Timefield** - Number of seconds passed since 00:00:00 of April 27th, 2015.
- **Latitude** and **Longitude** - Set of coordinates which characterize the geographical localization of the site, according to the World Geodetic System (WGS84).
- **CellId1** to **CellId6** - Unique ID number which identifies a cell, corresponds to the **CELL_REF** in Chapter 3. Up to 6 cell measurements can be presented but the UE is always camped in cell with ID number specified in **CellId1**.
- **RSCP1** to **RSCP6** - RSCP measurements for each reported cell.
- **SC1** to **SC6** - Scrambling Code being used by the UE in each cell.
- **ECNO1** to **ECNO6** - $E_c/N_0$ (carrier power to noise ratio) measurement for each reported cell.
These traces allow operators to have a very detailed picture of their networks, by knowing the signal strength of each user, in each moment and in the site. It is even possible to know the journeys each user made during the sample period. With this information a lot of conclusions can be taken, such as the existence of coverage holes or load imbalances between frequency layers. Therefore these traces are extremely valuable in developing real-time optimizing algorithms such as the one being developed in this thesis.

Data Parsing

The available data was stored in an Excel Worksheet File so it is necessary to treat it in the same way as in Chapter 3, loading this file and saving the data into a Table type variable and storing it in a MAT file for easy access. Afterwards, it is necessary to filter the data since there is some information regarding cells not used in Chapter 3. Due to this constraint, it was decided to organize the data in a structure containing the site number for easier access to relevant data in each function.

4.3 Imbalance Analysis

Before applying the load balancing algorithm to a site, it is beneficial to know the degree of imbalance between the 900 MHz and 2100 MHz frequency bands present in the site, which will be referred to as U900 and U2100 from here onwards. For this purpose, a small routine was designed.

Firstly, it evaluates two factors: the percentage of capacity used by each band and the ratio between the allocated capacity in each band. Then it checks how far from perfect balance the site is. This is done by comparing the imbalance factor of used capacity ($I_{usage}$) and the imbalance factor of the allocated capacity ($I_{capacity}$), originating the imbalance differential ($\Delta I$). Ideally this differential should be equal to zero, but most times it is bigger since the U2100 band is usually overloaded, as a consequence of what was explained in Section 2.1.3.

The 3 factors are detailed in Equations 4.1, 4.2 and 4.3, where $n_{CEU900}$ and $n_{CEU2100}$ represent the average CE usage reported on the corresponding day for both frequency bands and
$n_{CE,limU2100}$ and $n_{CE,limU900}$ represent the allocated capacity in both frequency bands. These values were taken from the traffic statistics file used in Chapter 3 for the forecasting algorithm.

\[
I_{usage} = \frac{n_{CE,U900}}{n_{CE,U2100} + n_{CE,U900}} \times 100[\%] \quad (4.1)
\]

\[
I_{capacity} = \frac{n_{CE,limU900}}{n_{CE,limU2100} + n_{CE,limU900}} \times 100[\%] \quad (4.2)
\]

\[
\Delta I = I_{capacity} - I_{usage} \quad (4.3)
\]

Results and Analysis

The imbalance factors were calculated for all 43 sites in order to determine which sites could be improved with the Load Balancing algorithm. The results are displayed in the graph of Figure 4.1, where the blue bars represent the $\Delta I$ for each site and the red line represents the decision threshold to select the candidate sites for Load Balancing.

![Imbalance Differential](image.png)

*Figure 4.1: Imbalance factors for all sites.*
It was decided that a site must have an imbalance differential larger than 10% to be a candidate for improvement. This means that a site is overloaded in the U2100 band while there is still considerable capacity available on the U900 band. From the 43 initial sites 24 were selected, but Site 35 was later removed due to lack of available data. There are a few sites with negative imbalance differential, only one with a factor greater than 10%, but these sites will not be considered since the objective of this thesis is to offload the U2100 band. The selected sites are presented in Table 4.2 along with the corresponding imbalance factors.

<table>
<thead>
<tr>
<th>Site</th>
<th>( I_{\text{capacity}} ) [%]</th>
<th>( I_{\text{usage}} ) [%]</th>
<th>( \Delta I ) [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41.46</td>
<td>21.59</td>
<td>19.87</td>
</tr>
<tr>
<td>4</td>
<td>32.76</td>
<td>18.85</td>
<td>13.91</td>
</tr>
<tr>
<td>5</td>
<td>23.81</td>
<td>9.23</td>
<td>14.58</td>
</tr>
<tr>
<td>7</td>
<td>38.46</td>
<td>24.23</td>
<td>14.23</td>
</tr>
<tr>
<td>9</td>
<td>33.33</td>
<td>18.81</td>
<td>14.52</td>
</tr>
<tr>
<td>10</td>
<td>21.43</td>
<td>10.91</td>
<td>10.51</td>
</tr>
<tr>
<td>18</td>
<td>34.48</td>
<td>18.22</td>
<td>16.27</td>
</tr>
<tr>
<td>19</td>
<td>37.84</td>
<td>26.04</td>
<td>11.80</td>
</tr>
<tr>
<td>20</td>
<td>57.14</td>
<td>36.26</td>
<td>20.89</td>
</tr>
<tr>
<td>21</td>
<td>31.91</td>
<td>17.12</td>
<td>14.79</td>
</tr>
<tr>
<td>23</td>
<td>43.86</td>
<td>30.19</td>
<td>13.67</td>
</tr>
<tr>
<td>24</td>
<td>50.00</td>
<td>30.35</td>
<td>19.65</td>
</tr>
<tr>
<td>25</td>
<td>30.00</td>
<td>13.48</td>
<td>16.52</td>
</tr>
<tr>
<td>26</td>
<td>46.67</td>
<td>30.18</td>
<td>16.49</td>
</tr>
<tr>
<td>27</td>
<td>37.14</td>
<td>18.69</td>
<td>18.46</td>
</tr>
<tr>
<td>28</td>
<td>42.86</td>
<td>15.35</td>
<td>27.51</td>
</tr>
<tr>
<td>34</td>
<td>41.03</td>
<td>21.95</td>
<td>19.07</td>
</tr>
<tr>
<td>36</td>
<td>38.46</td>
<td>17.41</td>
<td>21.05</td>
</tr>
<tr>
<td>37</td>
<td>33.33</td>
<td>21.69</td>
<td>11.65</td>
</tr>
<tr>
<td>39</td>
<td>36.36</td>
<td>21.11</td>
<td>15.25</td>
</tr>
<tr>
<td>40</td>
<td>42.86</td>
<td>22.92</td>
<td>19.94</td>
</tr>
<tr>
<td>42</td>
<td>42.86</td>
<td>16.79</td>
<td>26.06</td>
</tr>
<tr>
<td>43</td>
<td>52.00</td>
<td>20.05</td>
<td>31.95</td>
</tr>
</tbody>
</table>

In Figure 4.2 the Cumulative Distribution Function (CDF) of the imbalance differential \( (\Delta I) \) is presented. Through this graph we can understand that a large number of sites is incorrectly balanced and can be improved with the Load Balancing algorithm. In fact, only about 45% of the total sites seem to have an acceptable level of balancing between the two frequency layers.
4.4 Load Balancing Algorithm

As previously mentioned, this algorithm is based on the idea of defining adaptable thresholds explored in [12], depending on the load situation of the sites. The Load Balancing algorithm was developed using MATLAB®, having as input the available traces from the selected sites and returning as output an optimized admission RSCP threshold for the U2100 band. Firstly, it is necessary to estimate traces for the U900 band, since data from this band is not available. After having this estimate it is possible to find the optimal RSCP admission threshold so that we obtain a close to perfect load balance, by analysing the imbalance factors of each site.

4.4.1 Parameter Estimation for the U900 Band

As mentioned earlier the trace data is not available for the U900 band, so it becomes necessary to estimate it. Since the objective of the Load Balancing algorithm will be to offload the U2100 band, it is not necessary to know exactly where the trace data is generated but only the number of events that are expected to occur in the U900 band.

To estimate this value we take the usage imbalance factor obtained earlier and calculate the expected number of events considering that it is proportional to CE usage. This is, of course,
assuming that the used capacity depends only on the number of events recorded in each band and that each event has the same weight in the total CE usage of the site. The estimate for this parameter is then given by Equation 4.4, where \( N_{\text{events}} \) represents the number of reported events in each band.

\[
N_{\text{events}_{U900}} = N_{\text{events}_{U2100}} \frac{I_{\text{usage}}}{100 - I_{\text{usage}}} \tag{4.4}
\]

### 4.4.2 Load Balancing Strategy

In the following sections, the strategy taken to perform the load balancing for the selected sites will be detailed. Firstly, the site is analysed to check if it is a viable candidate or not, this stage is explained in Section 4.3. Afterwards, a target for the number of events in both bands is calculated based on the imbalance factors. Finally, the required threshold to achieve the target load distribution is calculated.

After having a threshold estimate, the new distribution of the load can be made. This stage will be important to obtain an objective evaluation of the effectiveness of the Load Balancing algorithm.

### Target Number of Events

The target number of events is calculated based on the capacity imbalance factor, since this value is the optimal load distribution for each site. Taking the total CE usage and multiplying by the imbalance factor we obtain the target usage for the U900 band. The target number of events for the U900 band is rounded down to the closest integer in order to eliminate fractions of events from the calculations. Afterwards, the U2100 band usage is simply the remaining from the total usage. This process is detailed in Equations 4.5, 4.6 and 4.7.

\[
N_{\text{events}_{\text{total}}} = N_{\text{events}_{U2100}} + N_{\text{events}_{U900}} \tag{4.5}
\]
\[ N_{\text{target}_{U900}} = \left\lceil \frac{N_{\text{events}_{\text{total}}} \cdot I_{\text{capacity}}}{100} \right\rceil \]  
\( (4.6) \)

\[ N_{\text{target}_{U2100}} = N_{\text{events}_{\text{total}}} - N_{\text{target}_{U900}} \]  
\( (4.7) \)

**RSCP Threshold**

Originally, all cells have an admission threshold of -115 dBm. This means that only UEs with a higher or equal RSCP value may connect to the cell. The output of this Load Balancing algorithm will be a suggested RSCP threshold value in order to achieve a better load distribution between the two frequency bands.

Firstly, the vector containing the U2100 band RSCP values \((\text{RSCP})\) is sorted in ascending order. Afterwards, the RSCP value of the index \(n - N_{\text{target}_{U2100}}\) is stored as the new threshold, where \(n\) is the size of the RSCP vector. This is illustrated by Equation \(4.8\). Note that \(n = N_{\text{events}_{U2100}}\), which means it is the number of events in the U2100 band before the Load Balancing is applied.

\[ T_{\text{RSCP}} = \text{RSCP}_i, \quad i = n - N_{\text{target}_{U2100}} \]  
\( (4.8) \)

**New Load Distribution**

After having a suggested threshold, the number of events in each band \((N_{\text{result}})\) is calculated, as well as the new imbalance factor \((I_{\text{result}})\) obtained in order to understand what is the impact of the algorithm in the site capacity.

To calculate these parameters we have to check how many events have a reported RSCP value equal or superior to the obtained threshold for the U2100 band \((\text{RSCP}_{\text{event}} \geq T_{\text{RSCP}})\). These events will remain in the U2100 band, whereas the rest will now switch to the U900 band. The new number of U2100 events will be designated as \(N_{\text{result}_{U2100}}\). Equation \(4.9\) shows the approach taken to find the number of U900 events.
Having the resulting number of events for both bands, it is possible to calculate the new imbalance factor. The new imbalance differential ($\Delta I_{\text{result}}$) is also relevant to understand the effectiveness of the algorithm. Equations 4.10 and 4.11 detail how these factors are calculated.

$$I_{\text{result}} = \frac{N_{\text{resultU/900}}}{N_{\text{resultU/900}} + N_{\text{resultU/2100}}} \times 100\%$$ (4.10)

$$\Delta I_{\text{result}} = I_{\text{capacity}} - I_{\text{result}}$$ (4.11)

If the algorithm is to be considered effective, then the new imbalance factor should be close to the capacity imbalance factor, thus obtaining the perfect balance. This means that the imbalance differential should be close to zero. The analysis of the factor $\Delta I_{\text{result}}$ will then be the validation method for the Load Balancing algorithm. In Figure 4.3 the whole process is illustrated through a flowchart.

### 4.4.3 Results and Analysis

As mentioned in Section 4.4.2, the validation to the Load Balancing algorithm will be the $\Delta I$ parameter. In Table 4.3 the obtained values for $I_{\text{result}}$ and corresponding $\Delta I_{\text{result}}$ are presented, as well as the RSCP threshold values obtained in dBm for each selected site. In Figure 4.4 these results are illustrated next to the old imbalance values.

The results show a significant change in the $\Delta I$ parameter, which had values greater than 10% for the selected sites and now have an average value of 1.78%. This value demonstrates a good effectiveness for the Load Balancing algorithm since it comes close to the specified target of zero for the perfect balancing of each site.

However, Site 25 presents a value much higher than desired. This is due to a huge number of events with a low RSCP value, more specifically -115 dBm. This fact makes it very hard for
**Figure 4.3:** Flowchart of the Load Balancing algorithm.

**Figure 4.4:** Imbalance factors for all sites, before and after Load Balancing.
Table 4.3: Summary of Load Balancing results.

<table>
<thead>
<tr>
<th>Site</th>
<th>$I_{result}$ [%]</th>
<th>$\Delta I_{result}$ [%]</th>
<th>$T_{RSCP}$ [dBm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41.42</td>
<td>0.05</td>
<td>-98</td>
</tr>
<tr>
<td>4</td>
<td>32.65</td>
<td>0.11</td>
<td>-100</td>
</tr>
<tr>
<td>5</td>
<td>22.84</td>
<td>0.97</td>
<td>-104</td>
</tr>
<tr>
<td>7</td>
<td>37.90</td>
<td>0.56</td>
<td>-96</td>
</tr>
<tr>
<td>9</td>
<td>30.95</td>
<td>2.38</td>
<td>-103</td>
</tr>
<tr>
<td>10</td>
<td>21.39</td>
<td>0.04</td>
<td>-109</td>
</tr>
<tr>
<td>18</td>
<td>34.47</td>
<td>0.01</td>
<td>-104</td>
</tr>
<tr>
<td>19</td>
<td>36.68</td>
<td>1.15</td>
<td>-103</td>
</tr>
<tr>
<td>20</td>
<td>56.95</td>
<td>0.19</td>
<td>-96</td>
</tr>
<tr>
<td>21</td>
<td>30.93</td>
<td>0.99</td>
<td>-103</td>
</tr>
<tr>
<td>23</td>
<td>42.84</td>
<td>1.02</td>
<td>-100</td>
</tr>
<tr>
<td>24</td>
<td>48.48</td>
<td>1.52</td>
<td>-99</td>
</tr>
<tr>
<td>25</td>
<td>13.48</td>
<td>16.52</td>
<td>-115</td>
</tr>
<tr>
<td>26</td>
<td>45.41</td>
<td>0.01</td>
<td>-100</td>
</tr>
<tr>
<td>27</td>
<td>36.78</td>
<td>0.36</td>
<td>-99</td>
</tr>
<tr>
<td>28</td>
<td>41.93</td>
<td>0.93</td>
<td>-93</td>
</tr>
<tr>
<td>34</td>
<td>40.65</td>
<td>0.38</td>
<td>-105</td>
</tr>
<tr>
<td>36</td>
<td>37.50</td>
<td>0.96</td>
<td>-91</td>
</tr>
<tr>
<td>37</td>
<td>32.14</td>
<td>1.19</td>
<td>-102</td>
</tr>
<tr>
<td>39</td>
<td>29.63</td>
<td>6.73</td>
<td>-114</td>
</tr>
<tr>
<td>40</td>
<td>42.26</td>
<td>0.60</td>
<td>-105</td>
</tr>
<tr>
<td>42</td>
<td>40.00</td>
<td>2.86</td>
<td>-108</td>
</tr>
<tr>
<td>43</td>
<td>51.16</td>
<td>0.84</td>
<td>-96</td>
</tr>
</tbody>
</table>

the algorithm to decide on a suitable threshold for the site, since a higher value than the original -115 dBm will cause a capacity overload situation in the U900 band. Because of this, it decides to keep the original value unchanged. The slightly different values of $I_{result}$ and $\Delta I_{result}$ when comparing to Table 4.2 come from rounding errors.

In order to better understand how the Load Balancing algorithm affects the load distribution we can observe the cumulative distribution function one more time. In Figure 4.5 the new CDF and the old one are represented simultaneously. It can easily be concluded that after the Load Balancing about 95% of the sites show a good load distribution, opposing to the 45% presented earlier.
Coverage Impact

Reducing the RSCP thresholds of the sites has an impact on their coverage area, more specifically the coverage of the U2100 band. Theoretically, increasing the threshold will cause a reduction on the coverage area while decreasing it will cause the opposite effect, since the users further away from the site should have a lower RSCP due to the higher path loss they are subjected to.

In Figure 4.6 the coverage of Site 24 is illustrated, after the Load Balancing algorithm is applied. The blue dots represent events in the U2100 band and the red dots events that switched to the U900 band after changing the admission threshold. Despite what was predicted, the results do not show a strong correlation between switched events and their distance to the site. This happens because the reported RSCP value of a UE depends on other factors such as being indoor or outdoor. Indoor events tend to have lower RSCP values, due to penetration losses, when compared with outdoor events at the same distance to the site. This causes the appearance of red dots closer to the sites and an overall homogeneous looking coverage map. The coverage of Site 20 is illustrated in Figure 4.7. In this case, the reduction of coverage of the U2100 band is more perceptible.
Figure 4.6: Coverage of Site 24 after Load Balancing, in blue are represented events which remained in the U2100 band and in red are represented events which switched to the U900 band.

Figure 4.7: Coverage of Site 20 after Load Balancing, in blue are represented events which remained in the U2100 band and in red are represented events which switched to the U900 band.
Chapter 5

Capacity Limit Prediction

In this chapter both algorithms developed in Chapters 3 and 4 are combined to show the benefits of using the Load Balancing algorithm in the long-term picture of the Capacity Management of sites. Using the Traffic Forecasting algorithm it is possible to find an estimate of when each site will reach its capacity limit, before and after the use of Load Balancing.

5.1 Overview

The objective of this part of the work is to quantitatively determine the effect of the Load Balancing algorithm. Estimating when the capacity limit is reached before and after the Load Balancing and comparing the results it will be possible to determine the real gain. This gain can be presented according to several methods, such as site longevity or overall CE and economic savings.

5.2 Input Data

In this part of the work, the necessary input data will be the CE usage statistics used in Chapter 3 more specifically the UL usage since this will be the limiting factor for the capacity of the sites. However, since this data is only available before the use of Load Balancing, an estimate of the CE usage statistics after Load Balancing must be calculated.
The post Load Balancing estimate is based on the imbalance factors obtained, which are presented in Table 4.3. Taking the total CE usage in a site, i.e. summing the two frequency bands’ usage, and multiplying it by the imbalance factor an estimate of the U900 band CE usage is obtained. To get the same data for the U2100 band we simply have to subtract the previous value from the total. This process is demonstrated in Equations 5.1, 5.2 and 5.3 where \( n_{CE} \) is the vector containing the CE usage statistics before Load Balancing and \( n_{CE,LB} \) is the vector containing the estimates for CE usages after Load Balancing.

\[
n_{CE,total} = n_{CE,U900} + n_{CE,U2100} \tag{5.1}
\]

\[
n_{CE,LB,U900} = n_{CE,total} \Delta I_{result} \tag{5.2}
\]

\[
n_{CE,LB,U2100} = n_{CE,total} - n_{CE,LB,U900} \tag{5.3}
\]

In order to find an estimate of the capacity limit the full set of available data is used, spanning from November 28th of 2014 to June 4th of 2015, instead of dividing it into three parts, like it was done in Chapter 3. Moreover, since it produced lower prediction errors, a weekly conversion of the data is made, more specifically using the weekly 90\textsuperscript{th} Percentile.

An illustration of the input data used, weekly 90\textsuperscript{th} Percentile of CE usage in Site 2, is represented in Figure 5.1. In this case the Forecast algorithm skips the validation stage, and only executes the fitting and decision stages, using section 1 and 2 of the data respectively. Note that in this graph both baseband pools are summed, whereas in Figure 3.3 only one baseband pool is represented, thus the value discrepancies observed.

### 5.3 Improved Forecasting

In this chapter, since the input data for the Forecast algorithm is known in detail, it is possible to improve the algorithm in order to increase its accuracy. This improvement is done by dismissing one or several fitting functions, since they produce greater prediction errors.
In order to decide which functions are to be removed, the Forecast algorithm is ran for all U2100 Nodes and for the designated input data, i.e. weekly 90th percentile of the maximum UL CE usage, and the results are analysed. In Table 5.1 the results are grouped by fitting function chosen by the algorithm. It is evident that the Quadratic function produces the worse results in terms of prediction error. Due to this fact, this function will be removed from the Forecast algorithm in this chapter.

<table>
<thead>
<tr>
<th>Function</th>
<th>Average Error [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>2.45</td>
</tr>
<tr>
<td>Quadratic</td>
<td>15.79</td>
</tr>
<tr>
<td>Power</td>
<td>7.75</td>
</tr>
<tr>
<td>Gaussian</td>
<td>4.01</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>7.19</td>
</tr>
<tr>
<td>All</td>
<td>7.69</td>
</tr>
</tbody>
</table>

Afterwards, the Forecast algorithm is ran again only with the designated input data and it is validated by using the same method described in Chapter 3. The output is once again the prediction error, calculated with MAPE, for each site. Moreover, a sequential analysis is done by splitting the validation window into three equal parts (28 days each) and calculating the prediction for each section. This allows a better perception of how the error behaves with time.
In Table 5.2 the prediction error averaged over all the analysed sites is presented. It is easy to conclude this approach lowered the prediction error obtained in about 2%. The worst case and best case scenarios are also calculated, using a confidence interval based on the obtained standard deviation. Considering that the prediction error follows a normal distribution, then adding or subtracting the standard deviation twice, we obtain the upper and lower bounds respectively. This is detailed in Equation 5.4 where \( P \) is the probability of a given random variable \( x \) being in the specified region, \( \mu \) represents the mean value of the variable and \( \sigma \) its standard deviation.

\[
P\left(\mu - 2\sigma \leq x \leq \mu + 2\sigma\right) \approx 0.9545 \tag{5.4}
\]

The actual distribution of the error is, in fact, quasi-normal. In Figure 5.2 this distribution is presented. As can be seen in the histogram, most of the prediction error values, 84% to be more precise, are concentrated in the interval of one standard variation. About 95% of the prediction error values are, in fact, concentrated in the interval of two standard deviations, thus confirming the assumption made earlier.

Based on the results obtained in Table 5.2 it is possible to extrapolate the error obtained in predictions made farther in the future. In Figure 5.3 this is detailed. Day 0 is considered to be the first day in zone 3 of Figure 3.3, so between Day 0 and 56 the values presented are the values obtained in Table 5.2 and from there on is illustrated an expectation of the prediction error for the next 390 days using a linear regression. At the end of these 390 days, expected error is almost 40%, and this progression will be considered when calculating the longevity of the sites in the following section.
5.4 Results and Analysis

Using both cases, before and after Load Balancing, as input for the Forecast algorithm it is possible to obtain an estimate of when the capacity limit will be reached in a site. In Figures 5.4 and 5.5 is an example of how the Load Balancing algorithm affects the capacity limit of sites. In the first figure is illustrated the forecast obtained for the U2100 band of Site 21 before the Load Balancing process was applied, whereas in the second one the forecast obtained after the Load Balancing is considered. It is clear that on the first case, the site is on the verge of reaching its capacity limit, but after the Load Balancing algorithm is applied this instant is delayed for a
considerable amount of time.

**Figure 5.4:** Example of Forecast before Load Balancing.

**Figure 5.5:** Example of Forecast after Load Balancing.

The output of the algorithm is presented in Table 5.3 in terms of site longevity. This means that for each site, the algorithm returns the number of days until reaching the capacity limit. Since the algorithm was only ran for 390 days after the last available date some sites do not reach their capacity limit. These sites present a value of 390+ days, meaning that their
capacity limit will be reached after more than 390 days. In the table, the sites which were subjected to the load balancing algorithm are coloured in light grey.

In the same table are also considered the best and worst case scenarios, using the errors obtained in Figure 5.3 through linear regression. The values presented were calculated considering that the capacity limit might be above or under the mean value by as much as the expected error of the corresponding month.

To better understand the impact of the results, the corresponding complementary CDF was also calculated. It is presented in Figure 5.6. The graph shows that, before Load Balancing, 80% of the sites analysed have a predicted longevity of more than 3 months, whereas after the Load Balancing process the predicted longevity is greater than 1 year for the same amount of sites. Another way of interpreting this graph is realizing that, before load balancing, only about 50% of the sites have a longevity greater than one year. After the Load Balancing this figure rises to 80%.

![Capacity Limit complementary CDF](image)

**Figure 5.6: Capacity Limit complementary CDF.**

However, a more practical approach to look at this issue is to estimate the amount of CEs that the Load Balancing process allows to save after a certain amount of time. Figure 5.7 illustrates the amount of CEs needed for expansion in all sites for a time duration of up to one year. It also includes a confidence interval for the forecast, obtained by considering the mean value of the expected prediction error for each month. The dotted lines represent this interval for both cases.
Figure 5.7: CEs needed for expansion of all sites over one year.

It is possible to verify that in one year, before Load Balancing, the operator would need to expand the sites in about 1400 CEs. After applying the Load Balancing this value drops to around 400 CEs, representing a gain of about 1000 CEs. In terms of time this would save the operator between 6 and 7 months before having to make an expansion of just 400 CEs. This gain increases exponentially with time.
<table>
<thead>
<tr>
<th>Site</th>
<th>Before Load Balancing</th>
<th>After Load Balancing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Worst Case</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>77</td>
<td>91</td>
</tr>
<tr>
<td>2</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>3</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>77</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>9</td>
<td>71</td>
<td>84</td>
</tr>
<tr>
<td>10</td>
<td>192</td>
<td>280</td>
</tr>
<tr>
<td>11</td>
<td>169</td>
<td>231</td>
</tr>
<tr>
<td>12</td>
<td>233</td>
<td>378</td>
</tr>
<tr>
<td>13</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>14</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>15</td>
<td>86</td>
<td>105</td>
</tr>
<tr>
<td>16</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>17</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>18</td>
<td>125</td>
<td>168</td>
</tr>
<tr>
<td>19</td>
<td>65</td>
<td>77</td>
</tr>
<tr>
<td>20</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>21</td>
<td>71</td>
<td>84</td>
</tr>
<tr>
<td>22</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>23</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>24</td>
<td>174</td>
<td>245</td>
</tr>
<tr>
<td>25</td>
<td>146</td>
<td>196</td>
</tr>
<tr>
<td>26</td>
<td>125</td>
<td>168</td>
</tr>
<tr>
<td>27</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>28</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>29</td>
<td>197</td>
<td>287</td>
</tr>
<tr>
<td>30</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>31</td>
<td>92</td>
<td>112</td>
</tr>
<tr>
<td>32</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>33</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>34</td>
<td>187</td>
<td>273</td>
</tr>
<tr>
<td>35</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>36</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>37</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>38</td>
<td>233</td>
<td>378</td>
</tr>
<tr>
<td>39</td>
<td>230</td>
<td>371</td>
</tr>
<tr>
<td>40</td>
<td>125</td>
<td>161</td>
</tr>
<tr>
<td>41</td>
<td>390+</td>
<td>390+</td>
</tr>
<tr>
<td>42</td>
<td>98</td>
<td>119</td>
</tr>
<tr>
<td>43</td>
<td>92</td>
<td>112</td>
</tr>
</tbody>
</table>
Chapter 6

Conclusions

This chapter brings a conclusion to the thesis, presenting a summary of the outcomes reached during this project as well as possible future improvements that may be added following its results.

6.1 Summary

During the course of this thesis, a solution regarding the optimization of Capacity Management in 3G Wireless Access Networks was reached, using a Load Balancing algorithm and evaluating its impact by using a Traffic Forecast algorithm. This work was developed using real traffic statistics and resource usage in sites that present two co-located frequency bands, where the U2100 band is usually overloaded when compared to the U900 band. This data was provided by a Portuguese mobile telecommunications operator. Moreover, this thesis was developed in collaboration with Celfinet, a Portuguese telecommunications consulting company. This project presents a solution for real-time optimization of the network based on real geo-positioned indicators, contributing to the development of Self-Organizing Networks.

Firstly, the Forecast algorithm was developed and validated by calculating the prediction error associated with the forecasts obtained. Using several types of input data it was also possible to conclude that the best approach to take was to predict the weekly progression of traffic
statistics rather than its daily behaviour. Another feature developed was a weekly characteri-
ization of the available sites, which would allow classifying sites into weekend or working days
sites according to their usage. This would allow having even lower prediction errors, but since
the sites available for study did not present a great diversity, this approach was ultimately not
implemented in the Forecast algorithm.

Afterwards, the Load Balancing algorithm was developed. It analyses imbalance factors
to decide which sites have an improper load distribution and later tries to distribute the load
equally between the two frequency bands by changing the admission thresholds of the U2100
band. After having a suggested threshold, the imbalance factors were once again calculated.
Analysing them, it was concluded that using the Load Balancing algorithm it was possible to
raise the number of correctly balanced sites from 45% to 95% of the total set of sites.

Finally, in Chapter 5, the capacity of the sites was studied in order to find out when its
limit is reached. Firstly, the limiting factor was defined as being the maximum UL CE usage,
since DL usage is far from reaching its limit. Knowing this would be the input data for the
Forecast algorithm, it was improved by eliminating one of the fitting functions. Afterwards, the
algorithm is ran in both cases, before and after the Load Balancing algorithm is applied. This
time the output of the Forecast algorithm is the longevity of each site, i.e. the number of days
until the capacity limit will be reached. Comparing the results for both cases, it was concluded
that the amount of sites with longevity of at least one year is raised by about 30% and that
after a single year it is possible to save about 1000 CEs in capacity expansions, which means a
reduction of costs for the operator.

6.2 Future Improvements

In this section will be presented some suggestions for further improvements that could be made
in this area. These possible improvements are spread between the following categories:

- **Longer input data period** - Having traffic statistics from a longer period of time it
  is possible to make more accurate predictions. Network operators have a vast amount of
  traffic statistics they can use to have a better idea of the longevity of their sites.
• **Larger site sample** - Having a higher number of sites with traffic statistics may be helpful to define some traffic behaviour patterns depending factors such as location, seasonality or rare events. This would also help increase the accuracy of the Forecast algorithm by characterizing sites in several categories and having different prediction methods for each category.

• **Multi-technology extrapolation** - In the current setting of wireless access networks various technologies co-exist, namely 2G, 3G and 4G technologies. This means that the developed Load Balancing algorithm may be used for evenly distribute traffic among the different technologies and frequency bands available, allowing an even greater increase in the longevity of sites.

• **Dynamic thresholds** - This Load Balancing algorithm’s output is a suggested admission threshold calculated with only one hour of trace data. Having a real-time dynamic system, such as the current wireless access networks, the thresholds can be updated along the day, enabling a greater efficiency for the algorithm. For example, the algorithm can evaluate the traffic statistics in each hour and decide on a threshold for the following hour, or choose an even smaller update frequency.

• **Event differentiation** - By knowing exactly how each event impacts capacity usage of a site, it is possible to develop an even more efficient solution for the Load Balancing algorithm, as opposed to the solution obtained which considers that all events have the same impact.
Bibliography


Appendix A

Blocking Probability Models

Algorithms like Kaufman-Roberts provide a method of obtaining a multi-service blocking probability for a certain resource. Given a limited capacity for a single resource $C_r$, the blocking probability $P_{B_i}$ for service $i$ is given by Equation (A.1). The distribution $q$ is expressed in terms of the resource consumption of the service, $b_i$, and of the offered traffic intensity, $A_i$, (in Erlang). This is detailed in Equation (A.2)

$$P_{B_i} = \sum_{i=0}^{b_i-1} q(C_r - i) \quad (A.1)$$

$$q(j) = \frac{1}{j} \sum_{i=1}^{k} A_i b_i q(j - b_i) \quad (A.2)$$

As in [8], it is computationally more efficient to use a solution based on the fast Fourier transformation (FFT), as detailed in Equation (A.3). Using this method, the probability distribution obtained is shown in Equation (A.4)

$$G = IFFT \left( \prod_{i=1}^{K} FFT(p_i) \right) \quad (A.3)$$

$$P_{B_i} = \frac{\sum_{i=C_r-b_K+1}^{C_r} G(i)}{\sum_{i=0}^{C_r} G(i)} \quad (A.4)$$
The blocking probability for all services is presented in Equation A.5, where \( P_{B_{ni}} \) is the blocking probability for service \( i \) at the resource \( n \) and \( c_{ani} \) represents the connection attempts for service \( i \) at the resource \( n \).

\[
P_{B_n} = 100 \frac{\sum_i c_{ani} P_{B_{ni}}}{\sum_n c_{ani}} \tag{A.5}
\]

Here, \( c_{ani} \) is given by Equation A.6, where \( c_{ani} \) for \( n = 1 \) represents the call attempts attended at the first subsystem.

\[
c_{ani} = c_{a(n-1)i} (1 - P_{B_{n-1}}), \quad n > 1 \tag{A.6}
\]

Once the resource consumption has been characterized for the different services and the traffic demand is known for each one, the modelling algorithm can estimate the per-resource and overall accessibility. As said in 2.1.7, there are three main resource constraints in UMTS: CE capacity, SC capacity and Iub capacity. Combining these three elements the overall accessibility (\( A_{cc} \)) can be obtained, for the downlink in Equation A.7 and for the uplink in Equation A.8. Note that, for the uplink, SC capacity is not relevant.

\[
A_{ccDL} = (1 - P_{B_{IubDL}})(1 - P_{B_{CEDL}})(1 - P_{B_{SCDL}}) \tag{A.7}
\]

\[
A_{ccUL} = (1 - P_{B_{IubUL}})(1 - P_{B_{CEUL}}) \tag{A.8}
\]