TÉCNICO LISBOA

Implementation Analysis of Cloud Radio Access Network Architectures in Small Cells

Tiago João Castro Gonçalves Frazão Monteiro

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisor: Prof. Luís Manuel de Jesus Sousa Correia

Examination Committee

Chairperson: Prof. José Eduardo Charters Ribeiro da Cunha Sanguino

Supervisor: Prof. Luís Manuel de Jesus Sousa Correia

Member of Committee: Prof. António José Castelo Branco Rodrigues

Member of Committee: Eng. Ricardo Dinis

November 2016

To my loved ones

Acknowledgements

First of all, I would like to thank Prof. Luís M. Correia for the opportunity to work under his supervision and for the guidance he provided me throughout the year. I am very thankful for the weekly meetings that taught me not only the technical skills but also skills related to presentation, work ethics and professional attitude. The lessons learned were very useful in the academic path, and will certainly be remembered in my professional life. I am also thankful for the opportunity to do this work in collaboration with industry, and to be a part of Group of Research On Wireless (GROW), where I had the right tools to finish the thesis.

To everyone in GROW for the way they received me and treated me during the development of the work. It was very beneficial to learn more about other fields of telecommunications and to enjoy the advice of other professionals. A special thanks goes to my friends and colleagues Hugo da Silva, José Guita, Benham Rouzbehani and Kenan Turbic, for all the work support and for the good moments.

To Eng. Ricardo Dinis, Eng. Pompeu Costa and Eng. Luis Santo from NOS, for providing me with the data and the lines of thought to use in my work. I am grateful for the time they had to follow my thesis and for the opportunity for this collaboration.

To all my close friends that I met in Instituto Superior Técnico and that have been with me throughout the years, and to all the other friends that have supported me in this journey.

To all of my family, especially my mother, Rosa Monteiro, father, Luís Monteiro, sister, Mariana Monteiro, and grandfather, Adelino Gonçalves, for their encouragement and support, and for providing me all the conditions to finish my work.

To my girlfriend, Maria Róis, for all the support, advice and comprehension that helped me throughout the last 5 years. Her friendship and love were very important during all this time.

Abstract

The main objective of this thesis was to analyse the performance of a Cloud Radio Access Network, in an already deployed LTE-A network, by taking advantage of the functional Remote Radio Head and Baseband Unit split, with the centralisation of the processing power in standard data centres. The work consisted of the analysis of the current macro base station location, and of a list of the possible data centres to present a possible C-RAN deployment under different strategies and algorithms that a mobile operator may be interested in. The metrics studied are fronthaul latency, pool capacity in traffic per hour, processing power capacity, and multiplexing gain. A total cost of ownership model is also presented to estimate cost savings possible with the technology. The model is implemented in a computational tool to provide a generic study of any scenario. The results obtained in the central area of Porto prove that C-RAN implementation with 19 pools is possible without introducing latency problems. In this kind of implementation, the operator can combine different traffic profiles, and achieve a multiplexing gain up to 1.31, which can be translated into capacity savings. A cost reduction of 63% in capital investment and 31% in operational expenditures is predicted for a centralised deployment, compared to another green field deployment without centralisation. A futuristic approach was also added to the approach, in order to simulate how massive small cells deployment to handle an 8-time increase in traffic affects results.

Keywords

LTE, LTE-A, Cloud RAN, Radio Access Network, RRH, BBU, Multiplexing Gain, Small cells.

Resumo

O principal objetivo desta tese foi analisar o desempenho de uma rede de acesso rádio implementada na *Cloud* no mesmo cenário de uma rede LTE-A já existente. O conceito consiste em dividir as estações base numa parte analógica (cabeça de rádio) e numa parte de processamento digital que é centralizada juntamente com as restantes unidades num centro de dados. O trabalho parte da estrutura atual da rede e de uma lista de possíveis centros de processamento, e analisa diferentes estratégias e algoritmos que poderão ter interesse para um operador móvel. As métricas que são estudadas dizem respeito à latência do *fronthaul*, à capacidade necessária em cada centro em tráfego por hora e em processamento, e ao ganho de multiplexagem. Um modelo de custos foi também desenvolvido para oferecer uma ferramenta que permita caracterizar quanto um operador poderá poupar. O modelo foi implementado em ferramentas computacionais para ser possível analisar automaticamente qualquer cenário. Da sua aplicação à área mais central do Porto, conclui-se que uma implementação com 19 centros de processamento não introduz problemas de latência. O maior ganho será o de multiplexagem devido à combinação de curvas diárias de tráfego, diferentes em várias células, resultando no valor de 1.31. Uma redução de 63% no investimento necessário e de 31% nas despesas anuais é prevista para este tipo de implementação. Adicionou-se também um algoritmo de proliferação para simular o crescimento de tráfego e uma possível rede baseada em células pequenas para suportar todos os pedidos dos utilizadores.

Palavras-chave

LTE, LTE-A, *Cloud* RAN, Rede de acesso rádio, cabeça de rádio, processamento banda base, ganho multiplexagem, células pequenas.

Table of Contents

Acknowledgements v
Abstract vii
Resumo viii
Table of Contents ix
List of Figures xii
List of Tables xv
List of Acronyms xvi
List of Symbols xix
List of Software xxiii
1 Introduction 1
1.1 Overview 2
1.2 Motivation and Contents 5
2 Fundamental Aspects 7
2.1 LTE Aspects 8
2.1.1 Network Architecture 8
2.1.2 Radio Interface 9
2.2 C-RAN and Virtualisation 12
2.2.1 Software-Defined Networks 12
2.2.2 Network Functions Virtualisation 14
2.2.3 Cloud Radio Access Network 16
2.3 Small Cells 19
2.4 State of the Art 21
3 Methodologies and Expected Results 25
3.1 Model Description 26
3.1.1 Problem Definition 26

3.1.2 Model Structure 26
3.2 Performance Parameters 28
3.2.1 Latency 28
3.2.2 Link Capacity 29
3.2.3 Processing Capacity 30
3.2.4 Multiplexing Gain 33
3.2.5 Dimensioning and Costs 33
3.3 Model Implementation and Algorithms 37
3.3.1 Generic Algorithm 37
3.3.2 RRH Proliferation Module 37
3.3.3 Minimise Delay Algorithm 40
3.3.4 Load Balancing Algorithm 40
3.3.5 Minimise Number of Pools Algorithm 42
3.3.6 Maximise Multiplexing Gain Algorithm 44
3.4 Model Assessment 45
4 Results Analysis 49
4.1 Scenarios 50
4.2 Reference Scenario Analysis 53
4.3 Latency Analysis 59
4.4 Uplink Analysis 60
4.5 Traffic Profiles Analysis 62
4.6 Number of Pools Analysis 64
4.7 Daytime Balancing Analysis 66
4.8 Proliferation Analysis 68
4.9 Cost Analysis 74
5 Conclusions 77
Annex A. Processing Power Reference Scenario 85
Annex B. BBU Pool Positioning 87
Annex C. Classification of cells 91
Annex D. Average Traffic Profiles 93
Annex E. Average Traffic Profiles 97
Annex F. Proliferation Scenarios 99
Annex G. User's Manual 103

References.....109

List of Figures

Figure 1.1. Voice and data traffic, 2010-2015 (extracted from [Eric15]). 2

Figure 1.2. Traffic and revenue decoupled (extracted from [Open10]). 3

Figure 1.3. OPEX and CAPEX example (extracted from [CMRI11]). 4

Figure 1.4. Mobile Cloud Architecture (extracted from [FPHG14]). 4

Figure 2.1. Overall EPS architecture of an LTE System (extracted from [HoTo11]). 8

Figure 2.2. Example of Resource Block allocation (extracted from [GeRK12]). 11

Figure 2.3. Major concepts of LTE transmission modes (extracted from [ADFJ09]). 12

Figure 2.4. SDN architecture with layers and interfaces (adapted from [JZHT14]). 14

Figure 2.5. Virtualisation of the EPC (extracted from [MSGB15]). 15

Figure 2.6. NFV architectural framework (extracted from [HGJL15]). 16

Figure 2.7. SDN, NFV and cloud computing relation (extracted from [MSGB15]). 16

Figure 2.8. C-RAN in an LTE network (extracted from [CCYS14]). 17

Figure 2.9. C-RAN architecture working with SDN and NFV (extracted from [Fuji14]). 19

Figure 2.10. C-RAN deployment scenarios (adapted from [CCYS14]). 21

Figure 3.1. Model structure. 27

Figure 3.2. Timing restrictions for LTE-A (extracted from [3GPP14]). 29

Figure 3.3. Model Implementation Algorithm. 38

Figure 3.4. Proliferation Module Algorithm 39

Figure 3.5. Minimise Delay Algorithm. 41

Figure 3.6. Load Balancing Algorithm. 42

Figure 3.7. Minimise Pools Algorithm. 43

Figure 3.8. Maximise Multiplexing Gain Algorithm. 44

Figure 3.9. Share of centralised cells depending on the maximum distance allowed. 46

Figure 3.10. Share of centralised cells depending on the capacity imposed in the network. 47

Figure 3.11. Number of sites depending on proliferation time. 47

Figure 3.12. Mean density of sites depending on proliferation time. 47
Figure 4.1. Porto scenario with sites and pools location. 50
Figure 4.2. Reference scenario with site classification. 53
Figure 4.3. Share of RRHs with FH distance (algorithm comparison). 54
Figure 4.4. Average capacity by each pool (algorithm comparison). 55
Figure 4.5. Multiplexing gain in each pool (algorithm comparison). 55
Figure 4.6. Computational processing power per pool (Load Balancing algorithm). 57
Figure 4.7. CAPEX and OPEX comparison (in 10 years). 57
Figure 4.8. CAPEX cost distribution. 58
Figure 4.9. OPEX cost distribution (10 years). 58
Figure 4.10. Share of traffic load being treated in a pool with maximum fronthaul distance. 59
Figure 4.11. Multiplexing gain with maximum fronthaul distance (Maximise MG algorithm). 59
Figure 4.12. TCO with maximum fronthaul distance (Maximise MG algorithm). 60
Figure 4.13. Average traffic in each pool within the day. 61
Figure 4.14. Average processing power in each pool within the day. 62
Figure 4.15. Average load per pool with different traffic profiles. 63
Figure 4.16. Average capacity in each pool. 63
Figure 4.17. Average two-way delay with the number of pools used. 64
Figure 4.18. Average capacity per pool in GBph with different scenarios. 65
Figure 4.19. Average capacity per pool in TOPS with different scenarios. 65
Figure 4.20. TCO with the number of pools in each scenario. 66
Figure 4.21. Overall network load during the day. 66
Figure 4.22. Number of pools required with maximum capacity per pool. 67
Figure 4.23. TCO of the network with capacity per pool. 68
Figure 4.24. Number of sites deployed with the number of proliferation years. 69
Figure 4.25. Maximum network load with the number of years of proliferation. 69
Figure 4.26. Maximum capacity in a pool with the time of proliferation. 70
Figure 4.27. Maximum processing capacity in a pool with the time of proliferation. 70
Figure 4.28. Number of pools required with the proliferation time. 71
Figure 4.29. Share of cell sites with each classification type. 72

Figure 4.30. Multiplexing Gain variation with shared of deployed RRHs. 73
Figure 4.31. TCO comparison with virtualisation in C-RAN. 74
Figure 4.32. TCO with different strategies in C-RAN. 75
Figure 4.33. Influence of cost factors on HW savings. 76
Figure 4.34. Influence of cost factor on OPEX savings. 76
Figure B.1. Pools identification (reference scenario). 88
Figure B.2. Pools identification (scenario with 1 pool). 88
Figure B.3. Pools identification (scenario with 4 pools). 88
Figure B.4. Pools identification (scenario with 8 pools). 89
Figure B.5. Pools identification (scenario with 12 pools). 89
Figure B.6. Pools identification (scenario with 16 pools). 90
Figure C.1. Algorithm for cell classification according to traffic. 92
Figure D.1. Average DL traffic variation in LTE (per cell). 94
Figure D.2. Average UL traffic variation in LTE (per cell). 94
Figure D.3. Average DL 3G traffic variation (per cell). 95
Figure D.4. Average DL 2G traffic variation (per cell). 95
Figure D.5. Average combined DL traffic variation (per cell). 95
Figure E.1. Normalised traffic profile used for each type of cell. 98
Figure F.1. Reference Scenario. 100
Figure F.2. Pessimistic Proliferation Scenario (traffic average strategy). 100
Figure F.3. Realistic Proliferation Scenario (traffic average strategy). 101
Figure F.4. Optimistic Proliferation Scenario (traffic average strategy). 101
Figure G.1. Example of input file configuration with traffic information. 106

List of Tables

Table 2.1. Relation between bandwidth, RB and carriers (adapted from [Corr15]). 11

Table 3.1. Typical data rates of CPRI links (adapted from [PCCR13]) 30

Table 3.2. Module Assessment Tests. 46

Table 4.1. Number of elements in the scenario. 51

Table 4.2. Reference values for cost layer. 52

Table 4.3. Scenario classification based on traffic profiles. 53

Table 4.4. Distance and latency values for reference scenario (algorithm comparison). ... 54

Table 4.5. Average capacity results (algorithm comparison) 55

Table 4.6. Multiplexing gain results based on traffic (algorithm comparison). 56

Table 4.7. Multiplexing gain types comparison. 56

Table 4.8. Average processing capacity results (algorithm comparison). 57

Table 4.9. Multiplexing gain results based on uplink traffic (algorithm comparison). 61

Table 4.10. Multiplexing gain comparison (maximise MG algorithm). 63

Table 4.11. Proliferation factor per year. 68

Table 4.12. Multiplexing Gains Achieved with different traffic models. 71

Table A.1. Processing power for physical layer (adapted from [DDLo15]). 86

Table A.2. Processing power for high layer functions (adapted from [DDLo15]). 86

List of Acronyms

1G	1st Generation
2G	2nd Generation
3G	3rd Generation
3GPP	3rd Generation Partnership Project
4G	4rh Generation
5G	5th Generation
ARQ	Automatic Repeat Query
AMC	Adaptive Modulation and Coding
API	Application Programming Interface
AuC	Authentication Centre
BBU	Baseband Unit
BPSK	Binary Phase-Shift Keying
BS	Base Station
CAPEX	Capital Expenditure
CDN	Content Distribution Network
CoMP	Coordinated Multipoint
CP	Cyclic Prefix
CPRI	Common Public Radio Interface
C-RAN	Cloud RAN
CS	Circuit Switch
DFT	Discrete Fourier Transform
DL	Downlink
DMTF	Distributed Management Taskforce
DU	Data Unit
eICIC	Enhanced Inter-Cell Interference Coordination
eNB	Evolved Node B
EPC	Evolved Packet Core
EPS	Evolved Packet System
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved Universal Terrestrial Radio Access
FDD	Frequency Division Duplex
FW	Firewall
GBR	Guaranteed Bit Rate
GPP	General Purpose Processor

GPRS	General Packet Radio Service
GTP	GPRS Tunnelling Protocol
HetNet	Heterogeneous Networks
HSS	Home Subscriber Server
ICIC	Inter-Cell Interference Coordination
IP	Internet Protocol
IMS	IP Multimedia Sub-System
ISG	Industry Specification Group
IT	Information Technology
LTE	Long Term Evolution
LTE-A	LTE Advanced
MBMS	Multimedia Broadcast and Multicast Services
MBR	Maximum Bit Rate
MG	Multiplexing Gain
MIMO	Multiple-Input and Multiple-Output
MME	Mobility Management Entity
MPLS	Multi-Protocol Label Switch
MU-MIMO	Multi-User MIMO
NAS	Non-Access Stratum
OFDM	Orthogonal Frequency Division Multiplex
OFDMA	Orthogonal Frequency Division Multiple Access
ONF	Open Networking Foundation
OPEX	Operational Expenditure
OTN	Optical Transport Network
PBCH	Physical Broadcast Channel
PCEF	Policy Control Enforcement Function
PCFICH	Physical Control Format Indicator Channel
PCRF	Policy and Charging Rules Function
PDCCH	Physical Downlink Control Channel
PDN	Public Data Network
PDSCH	Physical Downlink Shared Channel
P-GW	PDN Gateway
PHICH	Physical Hybrid ARQ Indicator Channel
PMIP	Proxy Mobile IP
PRACH	Physical Random Access Channel
PS	Packet Switch
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service

QPSK	Quadrature Phase-Shift Keying
RAN	Radio Access Network
RANaaS	Radio Access Network as a Service
RRH	Remote Radio Head
RoF	Radio Over Fibre
SAE	System Architecture Evolution
SC-FDMA	Single Carrier Frequency Division Multiplex
SDMA	Spatial Division Multiple Access
SDN	Software-Defined Network
SDR	Software-Defined Radio
S-GW	Serving Gateway
SIP	Session Initiation Protocol
SON	Self-Organising Network
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TFT	Traffic Flow Template
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
USIM	Universal Subscriber Identity Module
UTRAN	Universal Terrestrial Radio Access
VM	Virtual Machine
WAN	Wide Area Network
WDM	Wavelength Division Multiplex
WDM-PON	Wavelength Division Multiplex Passive Optical Network
WiMAX	Worldwide Interoperability for Microwave Access

List of Symbols

β	Added complexity for raw data latency switch
$\delta_{BBUpool}$	BBU pool processing delay
δ_{eNB}	Maximum delay budget for eNB
δ_{RRH}	Delay budget for RRH processing
δ_{RTT}	Round Trip Time for fronthaul transmission (two-way delay)
$\eta_{framing}$	Framing efficiency
ρ	Cost reduction of GPP
σ	Std. Deviation
η	Energy efficiency of GPPs compared to specialised hardware
θ	Added complexity in power suppliers
μ	Average
ν	Added complexity of baseline components
ω	Added complexity for low-latency switch
$A_{cabinet}$	Area of cabinets
A_{pools}	Area of pools
B	Bandwidth
B_{ref}	Reference scenario bandwidth
C_{alarm}	Cost of alarm unit
$C_{base/cell}$	Cost of baseline units per cell
C_{BBproc}	Cost of baseband processing board
$C_{cab/cell}$	Cost of cabinet per cell
C_{CAPEX}	Capital expenditure
$C_{control}$	Cost of control hardware
C_{CW}	Civil work costs
$C_{CW/cell}$	Civil work costs normalised per cell
C_{energy}	Cost of energy
C_{fan}	Cost of a fan unit
C_{FH}	Costs in fronthaul investment
$C_{FH/cell}$	Costs in fronthaul investment normalised per cell
C_{Gproc}	Cost of general processing board
C_{HW}	Hardware costs
$C_{HW,CRAN}$	Hardware cost in centralised deployment

$C_{HW,local}$	Hardware cost in decentralised deployment
$C_{IO/cell}$	Cost of interfaces per cell
$C_{IO,BH}$	Costs in backhaul interfaces
C_{IOFH}	Costs in fronthaul interfaces
C_{lic}	Costs due to licensing
$C_{lic/cell}$	Costs due to licensing normalised per cell
C_{line}	Line encoding ratio
$C_{mtc,local}$	Maintenance costs in local deployment
$C_{mtc,CRAN}$	Maintenance costs in C-RAN deployment
C_{OPEX}	Operational costs
C_{power}	Power consumption expenses
C_{Pconv}	Cost of power converter
C_{Psup}	Cost of full power supplier
C_{proc}	Cost of complete processing units
$C_{r,CRAN}$	Cost of square meter in centralised deployment
$C_{r,local}$	Cost of square meter in local deployment
$C_{rent,CRAN}$	Renting costs for centralised deployment
$C_{rent,local}$	Renting costs for local deployment
$C_{structure}$	Cost of physical structure
$C_{SW,l}$	Cost of low-latency switch
$C_{SW,r}$	Cost of raw data switch
C_T	Total cost of ownership
d_{FH}	Fronthaul allowed distance
E	Spectral efficiency
E_{ref}	Reference scenario spectral efficiency
F_{DC}	Frequency domain duty-cycle
$F_{DC,ref}$	Reference scenario frequency domain duty-cycle
f_s	Sampling Rate
G_{mux}	Overall multiplexing gain
$G_{mux,p}$	Multiplexing Gain in pool p
G_{proc}	Overall processing gain
G_{total}	Overall gain with centralisation and under-provisioning
k_p	Number of racks in pool p
l_x	Share of maintenance costs with CAPEX
N_A	Number of antennas
$N_{A,ref}$	Reference scenario number of antennas
N_{BBU}	Number of BBUs in a pool
N_b	Sample width

$N_{BH/r}$	Number of backhaul interfaces per rack
$N_{BBboards}$	Number of baseband processing boards
$N_{Gboards}$	Number of general processing boards
$N_{cabinet}$	Number of cabinets
$N_{FH/cab}$	Number of fronthaul interfaces per cabinet
$N_{FH/r}$	Number of fronthaul interfaces per rack
N_{MIMO}	Number of antennas per sector (MIMO)
N_{pools}	Total number of pools
N_{RRH}	Total number of RRHs
$N_{RRH,p}$	Number of RRHs in pool p
N_{sh}	Number of shelves
N_{sites}	Number of sites
$N_{streams}$	Number of transmission streams
$N_{streams,ref}$	Reference scenario number of streams
O_x	Reference Value for processing power function
$P_{BBU,n}$	Processing power of a single BBU
P_{BH}	Processing power for backhaul
$P_{control}$	Processing power for control functions
P_{HLP}	Processing power in high layer protocols
P_f^{coding}	Processing factor for coding functions
$P_f^{channel}$	Processing factor for channel estimation functions
P_f^{DPD}	Processing factor for pre-distortion function
$P_f^{equaliser}$	Processing factor for equaliser compensation functions
$P_f^{equalisation}$	Processing factor for equalisation functions
P_f^{filter}	Processing factor for filtering functions
P_f^{FFT}	Processing factor for FFT/IFFT
P_f^{mapp}	Processing factor for mapping and demapping functions
P_f^{MIMO}	Processing factor for MIMO functions
P_f^{OFDM}	Processing factor for OFDM functions
$P_f^{sampling}$	Processing factor for sampling functions
P_f^{sync}	Processing factor for synchronisation functions
P_f^{TD}	Processing factor for time domain estimation functions
P_{fixed}	Fixed processing power in a BBU pool
P_{phy}	Processing power in the physical layer
P_{pool}	Required processing capacity in a BBU pool
P_{board}	Power consumed by specialised board
Q	Quantisation bits

Q_{ref}	Reference scenario quantisation bits
R_{CPRI}	Data rate required in the CPRI
$T^{peak}{}_{RRH}$	Peak traffic in most loaded RRH
$T^{peak}{}_{RRH,i}$	Peak traffic in ith RRH
v	Transmission speed in the link
V_p	Virtualisation factor in pool p
y	Number of years

List of Software

Microsoft Word 2016	Text editor software
Microsoft Excel 2016	Calculation and graphical software
Microsoft Visio 2013	Drawing algorithms software
Google Maps	Geographical plotting tool
Matlab	Computing environment

Chapter 1

Introduction

This chapter gives a brief overview of the thesis. A contextualisation of the topic in the current mobile communications scenario is described in Section 1.1, as well as the motivations and the scope of the work. At the end of the chapter, the document structure description is provided in Section 1.2.

1.1 Overview

Mobile communications systems provide their users the possibility to communicate to each other wherever they are through their mobile equipment. Although the first systems were designed only for voice communication, data exchange has become the main source of traffic in mobile networks. Driven by the evolution of the internet and the increasing computational resources of the user's mobile equipment (smartphones and tablets), operators have been expanding and upgrading their networks to cope with the new traffic demands. The number of mobile subscribers has been globally increasing around 5% per year [Eric15]. Besides that, the number of broadband subscriptions is expected to be around 85% of all subscriptions by the end of 2020. The increasing number of smartphones, tablets and personal computers connected to a mobile network means that the average generated traffic per user is also growing. The aggregated traffic by 2020 is expected to be around 9 times greater than the one measured in 2015, with video data representing around 60% of that number [Eric15]. Figure 1.1 represents the evolution of data and voice traffics since 2010.

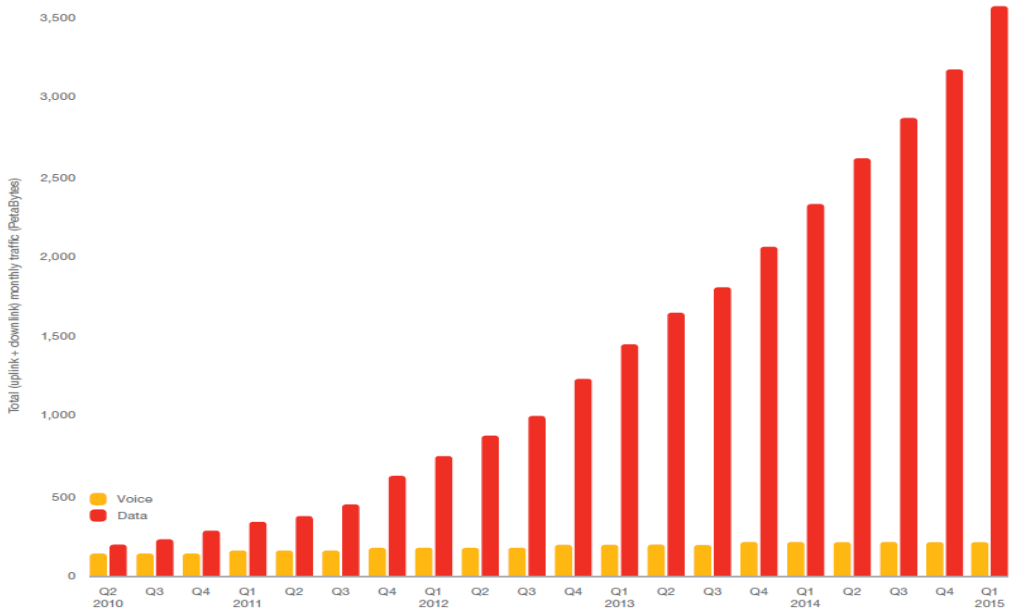

Figure 1.1. Voice and data traffic, 2010-2015 (extracted from [Eric15]).

The 1st Generation (1G) of mobile communications consisted of more than one independent analogue system. The Global System for Mobile Communications (GSM) was the first digital mobile communication system, being usually known as the 2nd Generation of mobile communications (2G). The later introduction of packet transmission with the development of General Packet Radio Services (GPRS) complemented with the radio interface improvements in GSM Evolution (EDGE), brought the possibility to do more than voice calls in a mobile terminal, allowing data services to be offered to end-users. The 3rd Generation Partnership Project (3GPP) is the organisation created to standardise the mobile communications systems after EDGE, mostly focused on providing improved capacity to end-users, while maintaining backward and forward compatibilities with others systems. Universal Mobile Telecommunications System (UMTS), one of the 3rd Generation (3G) of mobile communications systems, presents main upgrades relative to the previous system in the radio interface.

Long Term Evolution (LTE) is the 4th generation system (4G) and the one addressed in this thesis. The adoption of this technology was needed in order to improve data rates and reduce latency, according to the current traffic demands. In addition, reduced cost per bit, reasonable power consumption, simplified architecture with an all-IP network, spectrum efficiency, and compatibility with other systems were also pretended. The current version is known as LTE-Advanced (LTE-A), being standardised in Release 10 of 3GPP. Some of the enrichments introduced in this enhancement were Carrier Aggregation (CA), relaying, support for heterogeneous networks, and enhancements in technologies such as Multiple Input and Multiple Output (MIMO), interference control, coordinated transmission and self-organising networks (SON).

Mobile network operators have been upgrading their networks for years, with the intention to provide higher bandwidths to their customers and fulfil their traffic requirements. However, the revenue is not growing fast enough to ensure profitability. Figure 1.2 represents this tendency: not only traffic and revenues are decoupled, as revenue growth tends to stabilise or even to decline slowly.

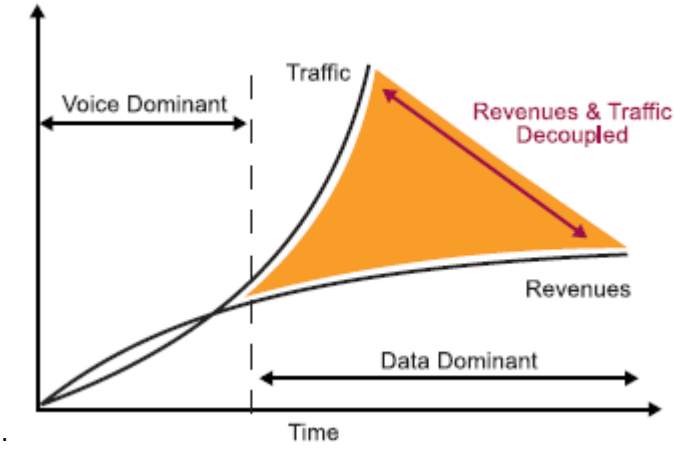

Figure 1.2. Traffic and revenue decoupled (extracted from [Open10]).

This situation creates new challenges to operators. The Total Cost of Ownership (TCO), which includes Capital (CAPEX) and Operational Expenditures (OPEX), is increasing in order to improve network infrastructures. In this way, network operators can satisfy their users and remain competitive between each other. Nevertheless, the Average Revenue per User (ARPU) is not growing significantly, as the typical mobile user is data-dependent, but still expects to pay less for data usage. With that said, operators need new architectures to optimise costs, while providing high-capacity to their subscribers. In addition, more efficient and greener solutions are expected due to energy-saving concerns.

It is usual to subdivide mobile network architectures into Radio Access Network (RAN) and Core Network (CN). The RAN is associated with the radio interface between the end-user equipment and Base Stations (BS). Besides that, it makes the connection of users to the CN. The RAN of LTE is addressed in this thesis. Figure 1.3 represents a study conducted by the China Mobile Research Institute to estimate the distribution of costs in a cell site (associated with CAPEX and OPEX).

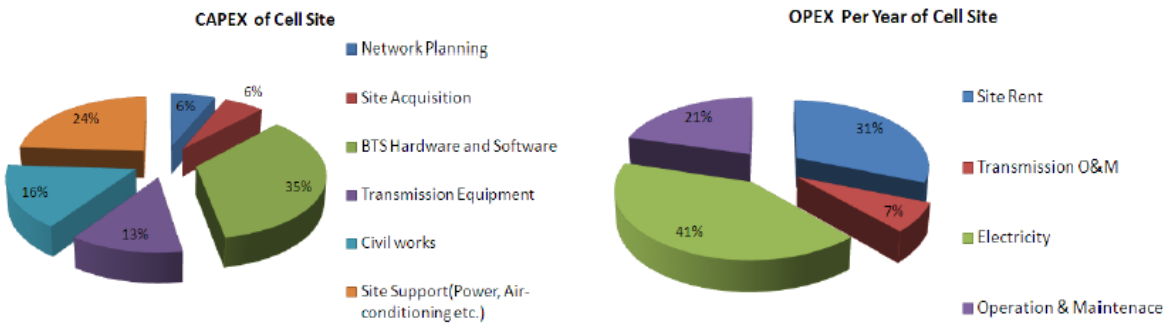

Figure 1.3. OPEX and CAPEX example (extracted from [CMRI11]).

Beyond the efforts to develop new solutions for the mobile network architecture to handle new challenges, new concepts to improve the overall performance of networks are being presented. Software-Defined Networks (SDN) is one of these concepts, intending to add an external controller with a global view of the network to efficiently manage all the elements of the network. In addition, cloud computing usage in IT environments has been attracting the attention of the telecommunications industry. Network Functions Virtualisation (NFV) comes as another concept willing to change the network, as its elements become virtual instances in data centres instead of the usual dedicated physical devices.

The concept of Cloud-RAN emerges as a solution of the mentioned problems for mobile operators. A centralised approach, in which all the signal processing needed for multiple base stations can be performed in one single physical location with increased computational resources, can reduce the global costs of the physical infrastructure, rent and operational expenditures in energy, Figure 1.3. Figure 1.4 represents a global view of this solution in which all the network is running over data centres. Savings in OPEX, such as electricity (mostly associated with air conditioning), as well as in CAPEX in civil works and transmission equipment appear as an advantage. Furthermore, upgradability and multi-standard support are facilitated with software-defined equipment in data centres. The concepts of C-RAN, SDN and NFV are highly complementary, and each one is expected to add value to the others.

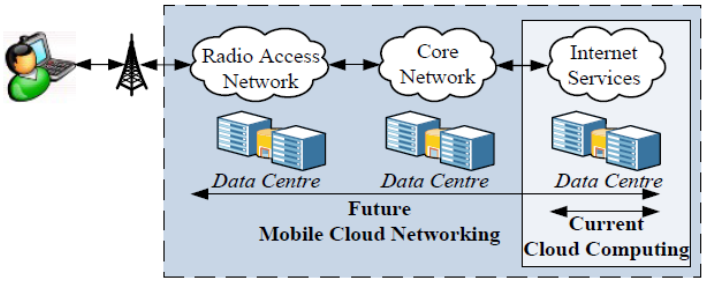

Figure 1.4. Mobile Cloud Architecture (extracted from [FPHG14]).

It is also important to note that such type of technology can easily be adapted to multiple networks. Besides the multi-standard support and the predicted software reconfigurability, centralised processing offers the possibility of tight coordination between base stations that are technological or geographically different from each other. Processing all the centralised data in powerful machines is quite promising

for the latest and future signal processing techniques and distributed algorithms, certainly improving aspects such as the achievable data rates and energy efficiency. With that being said, it is then expected to have such kind of deployment not only being considered for up-to-date networks but also studied for 5G networks, currently aiming at improvements in latency, data rates, energy consumption and connected devices by one or more orders of magnitude. Besides that, the prediction of future denser networks by deploying low-powered base stations can also take some benefits of this new approach.

1.2 Motivation and Contents

Cloud-RAN architectures have been seen as promising solutions to lower the costs associated with cell sites. Besides that, this kind of architectures might take full benefit of the new networks trends, such as SDN and NFV. When applied to LTE-A, C-RAN architectures also facilitate the implementation of advanced features, such as cooperative transmission and interference coordination, leading to improved throughput per user. The expected energy savings associated with C-RANs also allow classifying this solution as a “green” architecture.

Before mobile operators decide to use C-RANs, there are multiple issues that need to be addressed. Some of the problems are still related to the technologies required to implement C-RAN systems, such as fronthaul or baseband virtualisation. However, this work is focused on the problems associated with the implementation of C-RAN architectures in different network configurations. Taking the stringent requirements of latency, processing and transmission capacity of LTE-A into account, an analysis on how to deploy the C-RAN components in an already existent network and the performance achieved with that implementation is needed. Another important metric to quantify is the theoretical gain that can exist by centralising cells with different traffic profiles (residential of office traffic), and the implication of the value that may have in hardware reduction. A particular focus is given to heterogeneous networks architectures, in which the radio frequency components correspond to small cells. The goal of this last study is to propose an improved network with much more RRHs handling the new traffic demand that is expected in the following years. This layer is called proliferation of cells.

The purpose of this thesis was to develop a tool that implements a model to study the RRH-BBU assignments based on multiple criteria. The model was then applied to a dense urban scenario with already deployed LTE network, to analyse the performance of a migration to a C-RAN network. In the end, some of the main advantages of C-RAN are quantified. This study is done in collaboration with the Portuguese operator NOS, which provided the network configuration under study, the traffic curves and some of the assumptions used throughout the work.

The final output of this thesis is a computational tool that implements the model to analyse the performance parameters of a C-RAN deployment with base stations and pool locations, as well as traffic profiles as inputs. The tool is prepared to simulate multiple network configurations and multiple scenarios that can introduce benefits to mobile operators. A cost model was also developed to give the possibility

to compare the technical performance of the network with the possible cost savings of the C-RAN migration.

Regarding current works, the thesis adds up realistic traffic profiles of cells and their daily variation to previous thesis, such as [Maro15]. It also presents more realistic processing power data, as previous works were based on a less accurate model [WeGP13]. Besides, the thesis presents values for multiplexing gain such as [ChHC14], but with realistic values of traffic and including all type of cell profiles. It also compares the multiplexing gain of load with the gain observed with processing power. Finally, it also introduces some cost information to integrate the gain based on the operator viewpoint and some cost assumptions on operational expenditures to be considered.

The thesis is composed of 5 chapters and annexes. Chapter 1 is the present one, and provides an overview of the problem being solved and the motivation behind its study.

Chapter 2 starts with a brief description of LTE-A network architectures and radio interfaces. Then, a brief description of the SDN, NFV and C-RAN concepts is presented. Small cells are also described in the context of C-RAN architectures. The chapter concludes with the state of the art describing the most recent and relevant works on the subject of this thesis.

Chapter 3 presents the developed model and the parameters to be analysed. It stars by a brief description of the problem and the presentation of the performance parameters under evaluation in the thesis. Afterwards, the model structure and the algorithms developed are presented and explained in detail with the support of flowcharts and descriptions. Finally, the model is evaluated to ensure the rightness of the computational tool and the correct behaviour of the algorithms.

Chapter 4 presents the analysis of the obtained results and the scenarios description. It starts by defining the reference scenario and justifying the assumptions taken. Then, a study of the parameters in the reference scenario is done followed by different analysis where some parameters are changed to measure their impact on the network. The results are presented along the chapter together with the corresponding discussion.

Chapter 5 summarises the main results obtained in Chapter 4 and presents the final conclusions of this work. It also presents suggestions for future work that may complement the results obtained in this thesis.

In the end, a group of annexes is also presented that support some of the information presented in the main body of the thesis. Annex A presents the reference tables that are used in the power model. Annex B presents the scenario under study with the different number of possible pools. Annex C presents the algorithm that is used to classify cells according to their traffic type. Annex D presents the average curves of the scenario based on the classification of Annex C. Annex E presents the standard traffic profiles that are used in traffic proliferation models. Annex F presents the positioning of new cells in the proliferation layer. Annex G presents a user manual that is useful to run the computational tool developed under this work for other scenarios.

.

Chapter 2

Fundamental Concepts

This chapter provides an overview of LTE and Cloud Radio Access Networks. Section 2.1 focus on the main aspects regarding network architecture and radio interface of LTE. Section 2.2 addresses the characteristics of C-RAN architectures. Section 2.3 discusses small cells in LTE-A environments. Then, Section 2.4 is fully dedicated to an analysis of the state of the art relative to C-RAN in LTE.

2.1 LTE Aspects

2.1.1 Network Architecture

In this section, an overview of LTE network architecture is given, based on [DaPS11], [HoTo11] and [SeTB11].

LTE has been developed to support only Packet-Switched (PS) services in contrast to the circuit-switched (CS) approach of previous networks. The goal was to provide continuous connectivity between User Equipment (UE) and the Packet Data Network (PDN) through the Internet Protocol (IP). The Evolved Universal Terrestrial Radio Access (E-UTRAN) of LTE is the evolution of the UMTS radio access, and it is complemented by the improvement of other network features, which includes the Evolved Packet Core (EPC) network, known as the *System Architecture Evolution* (SAE). LTE and SAE are the two main elements of Evolved Packet System (EPS) standardised by 3GPP.

Figure 2.1 shows the overall network architecture covering the main elements and the interfaces between them. A dashed line represents a control-plane connection and a solid one a user plane connection. Four main domains constitute the network and are equivalent to those existing in previous systems: services, EPC, E-UTRAN and UE. The major improvements in LTE were done in E-UTRAN and in EPC.

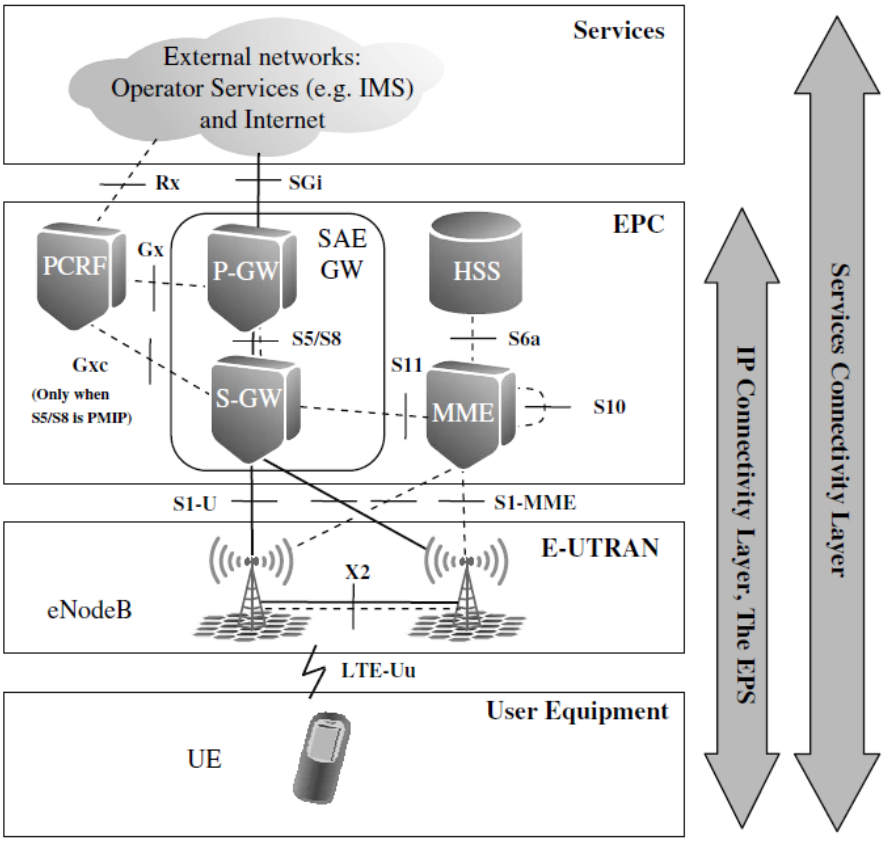

Figure 2.1. Overall EPS architecture of an LTE System (extracted from [HoTo11]).

EPS uses the concept of bearers, i.e., an IP packet flow with a defined Quality of Service (QoS). Multiple bearers can be established (even to different PDNs) and released by the E-UTRAN and EPC together as required by applications. As all services are provided on top of IP, IP Multimedia Sub-System (IMS) is an architectural framework used by operators in the Services Connectivity Layer to provide services, such as Voice over IP (VoIP) and interconnectivity to legacy CS networks. Security and privacy of the user and the network are also provided by the EPS.

The UE contains the Universal Subscriber Identity Module (USIM) placed in a removable smart card, used to identify and authenticate the user and to generate security keys.

E-UTRAN comprises only one node in a flat architecture called E-UTRAN Node B (eNodeB or eNB). The eNodeBs are typically distributed throughout the entire coverage area controlling one or more cells, handling all the radio communications between the UE and the EPC. The main functions of eNodeB are to provide security (ciphering and deciphering data in both links), offer Radio Resource Management (all functions related to radio bearers, for instance radio mobility and dynamic allocation of resources), perform header compression, assist the positioning service and deliver connectivity to the EPC, providing signalling and the bearer path towards to S-GW. Multiple S-GW/MMEs can be connected (S1 interfaces) in order to allow load sharing and redundancy (creating *pools*). The X2 interface connects different eNodeBs and it is mainly used to support UE mobility and multi-cell functions, as Inter-Cell Interference Coordination (ICIC), defining a distributed control system, unlike previous technologies.

The EPC is responsible for control of the UE and establishment of bearers, its main nodes are:

- PDN Gateway (P-GW) is connected to external PDNs, and it is the highest level mobility anchor in the network. It is responsible for IP allocation to the UE, as well as QoS assurance and charging according to directives from the PCRF. QoS enforcement is based on downlink filtering Traffic Flow Templates (TFT) to provide different Guaranteed Bit Rates (GBR) to different bearers. The P-GW is also the connection to non-3GPP radio access technologies, e.g. WiMAX.
- Serving Gateway (S-GW) transmits all IP packets. It serves as a local reference point for the data bearers when the UE moves among eNodeBs and for inter-working with other networks, such as GSM and UMTS. It also holds bearer information when the UE is idle and performs some administrative jobs in the visited network (e.g., data for charging and legal issues).
- Home Subscriber Service (HSS) contains users' data, such as the subscribed QoS profile and roaming restrictions. It also records the allowed PDN connections and user location at the level of visited network control element, such as MME. It can include the Authentication Centre (AuC).
- Policy and Charging Resource Function (PCRF) makes decisions on how to handle services in terms of QoS and provides information to the Policy Control Enforcement Function (PCEF) in P-GW, ensuring that the data flow is treated accordingly to users' subscription.
- Mobility Management Entity (MME) is an essential control plane node responsible for security functions (e.g., authentication and authorisation), handling idle state mobility, roaming and handovers. It is also responsible for the establishment, maintenance and release of bearers. The protocols between MME and UE are known as *Non-Access Stratum* (NAS).

2.1.2 Radio Interface

This section addresses the main characteristics of LTE's Radio Interface based on [HoTo11], [Corr15] and [SeTB11].

According to 3GPP Release 12 [3GPP15], LTE is designed to operate in 44 frequency bands, each standardised to operate in a specific duplex mode that can be either Time Division Duplex (TDD) or Frequency Division Duplex (FDD). In Portugal, ANACOM (the telecommunications regulator) conducted an auction in which the three operators chose to use the 800 MHz, 1 800 MHz and 2 600 MHz bands for LTE [ANAC11].

Regarding Downlink (DL) communication, a transmission based on Orthogonal Frequency Modulation (OFM) is deployed. The principle of this modulation is to use multiple subcarriers with the total data rate of the system being divided among all of them. With the concept of orthogonality between subcarriers, it is possible to remove the typical guard bands and to improve the channel spectral efficiency. The multiple access technology is the multiple access version of Orthogonal Frequency Division Multiplexing (OFDM) known as Orthogonal Frequency Division Multiple Access (OFDMA) and the concept is to consent different users to use different sets of subcarriers to send their content. A constant frequency spacing between subcarriers has been selected to be 15 kHz. A Cyclic Prefix (CP) is added to each symbol in order to avoid Inter-Symbol Interference (ISI). The CP length is variable according to the cell type. The main advantages brought by OFDMA are (extracted from [HoTo11]):

- Good performance in frequency selective fading channels.
- Low complexity of baseband receiver.
- Good spectral properties and handling of multiple bandwidths.
- Link adaptation and frequency domain scheduling.
- Compatibility with advanced receiver and antenna technologies.

The main disadvantages of this technology are the tolerance to frequency offsets and the high Peak-to-Average Ratio (PAR).

In the Uplink (UL), the PAR problem is harder to solve for mobile devices that are power-limited. A Single Carrier Frequency Multiplex Access (SC-FDMA) was implemented using DFT-spread techniques to mitigate the problem and achieve better power behaviour. The main aspects of this technology are similar to those described for ODFMA and a Cyclic Prefix is also used. The major difference is that, unlike OFDM where data symbols directly modulate each subcarrier, in SC-FDMA the signal modulated in a single subcarrier is a linear combination of all the data symbols transmitted at the same time.

Frames have a duration of 10 ms in both links. As LTE supports both FDD and TDD systems, two frame structures are possible. A frame is subdivided into ten 1 ms subframes, each of which is split into 0.5 ms slots. Each slot contains seven OFDM symbols with the normal CP length, or six if the extended CP is configured. The frequency domain resources are grouped into units of twelve subcarriers (in a total of 180 kHz), such that one unit for a duration of one slot is named as Resource Block (RB). There is still a smaller unit of resource named Resource Element (RE), which consist of only a subcarrier for a duration of one symbol. One RB comprises 84 REs in normal CP. Figure 2.2 is an example of resource allocation.

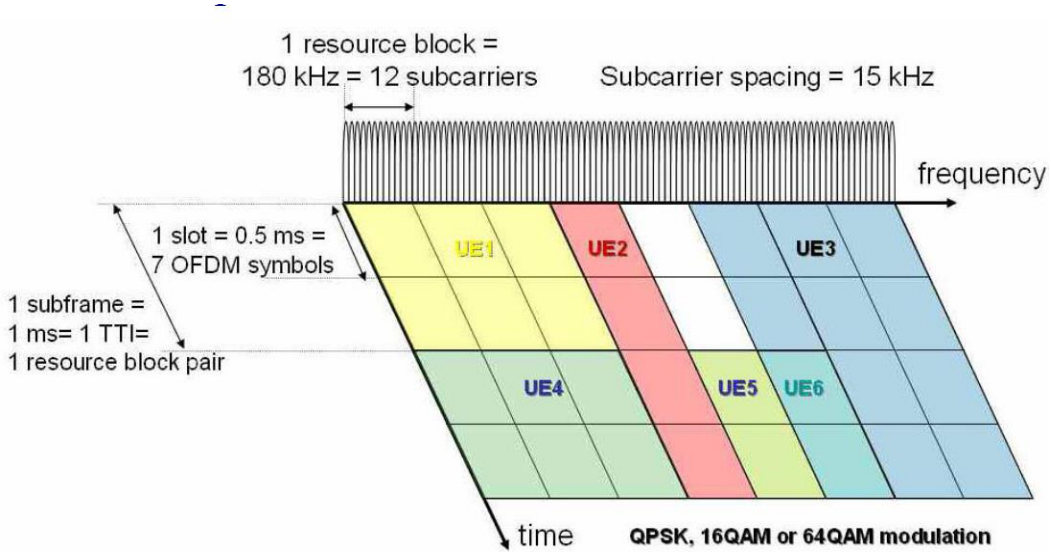

Figure 2.2. Example of Resource Block allocation (extracted from [GeRK12]).

The radio channel bandwidth can only assume the values presented in Table 2.1. The number of carriers and RBs available for each band is also presented.

Table 2.1. Relation between bandwidth, RB and carriers (adapted from [Corr15]).

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

LTE uses Adaptive Modulation and Coding (AMC) as well as adaptive transmission power, performing channel estimation achieved via pilot symbols or training sequences. QPSK, 16QAM or 64QAM (BPSK is also specified for control channels) and turbo code with different coding rates is used in order to provide the higher data rate possible to users over time-varying channels according to their specific situation within the cell.

Multiple-Input Multiple-Output (MIMO) is also an important enhancement of 4G systems compared to the previous Single-Input Single-Output (SISO) ones. A combined used of multiple antennas, either at the receiver or the transmitter sides (usually two or four antennas), is used to improve system capacity (more users per cell) and cell coverage (larger cells are possible), as well as service provisioning (higher data rates or higher signal-to-noise ratios). Multiple transmission modes have been standardised in successive 3GPP Releases around three major concepts. System Diversity offers robustness against multipath fading, by combining the same signal received through hopefully uncorrelated channels. Spatial Multiplexing is the transmission of more than one parallel streams to the same user on multiple spatial layers, providing higher data rates. Array Multiplexing is the use of array techniques to shape the overall beam of the antennas in a given direction, via precoding or beamforming (maximising the gain of the overall antenna). This last concept is also used to serve users in different locations simultaneously in time and frequency, the so-called Multi-User MIMO (MU-MIMO) or Spatial Division Multiple Access (SDMA). Coordinated Multipoint Transmission (CoMP) is another important feature of LTE-Advanced, consisting of coordinated geographically separated eNodeBs, providing jointly scheduling and jointly transmission. Figure 2.3 represents these concepts.

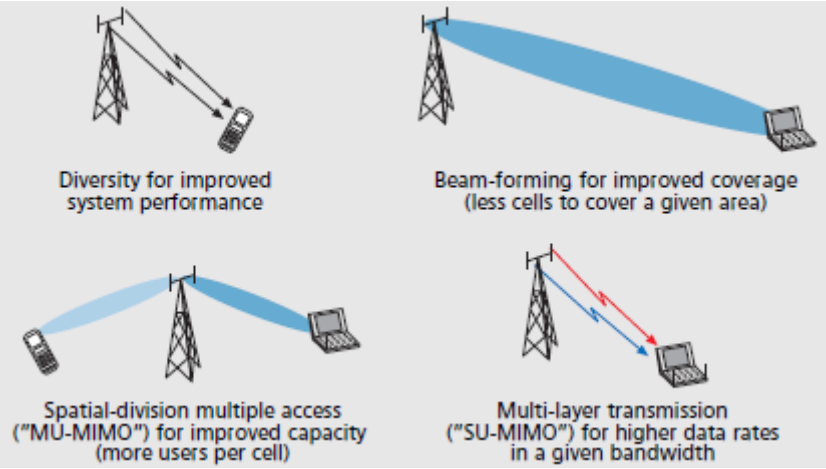

Figure 2.3. Major concepts of LTE transmission modes (extracted from [ADFJ09]).

Another enhancement introduced in LTE-A is Carrier Aggregation (CA), which is used to increase the bandwidth in a cell, and thus provide higher bit rates per user. Each aggregated carrier is referred to as a component carrier and their resource blocks are only used for data transmission. The aggregation can be performed contiguously by using adjacent frequency bands or non-contiguously using separated bands in frequency. In the latter case, the type of aggregation can also be classified according to the operating frequency bands used in each carrier, being the same or not (intra-bands or inter-bands).

There are 6 physical channels for DL. Physical DL Control Channel (PDCCH) for resource assignment, Physical Control Format Indicator Channel (PCFICH) for frame indication, and Physical Hybrid ARQ Indicator Channel (PHICH) for acknowledgments related to UL transmissions are the three channels that carry control information. Physical Broadcast Channel (PBCH) for vital parameters of the cell, Physical DL Shared Channel (PDSCH) for the main data and Physical Multicast Channel for Multimedia Broadcast and Multicast Services (MBMS) are the three transport channels. Three more channels exist in UL. The user content is transported on the Physical UL Shared Channel (PUSCH), the control data on the Physical UL Control Channel (PUCCH) and random access preambles sent to access the network are transported in the Physical Random Access Channel (PRACH).

2.2 C-RAN and Virtualisation

2.2.1 Software-Defined Networks

This section provides an overall perspective on Software-Defined Networks (SDN), explaining the main differences compared to traditional networks' architectures, the key principles and advantages, types of implementation, and the current challenges that need to be addressed.

The non-profit Open Networking Foundation (ONF) is the industry consortium leading the advancement of SDN and defines it as a network architecture in which the control and data planes are decoupled, network intelligence and state are logically centralised, and the underlying network infrastructure is

abstracted [ONFo13]. This kind of design has emerged mostly because of the rigid structure and the static state of the distributed networks that is no longer feasible for the current and future needs. Besides that, virtualisation trends being adopted by companies introduce an extra degree of complexity to the networks and a need to an efficient scalability of resources to answer to dynamic and unpredictable traffic patterns, leading to a time-consuming and expensive management (for instance, virtual machine migration requires network reconfigurations). Besides that, current vendors are using control-plane software to optimise their own data flows and achieve competitive advantages, making it harder to deploy worldwide network solutions [JSSA14]. Therefore, SDN stands as a solution that is efficient, flexible, agile and scalable.

The four key features regularly associated with SDN are [SSCF13]:

- Physical separation of the control plane from the data plane with the externalisation of the control plane to an outside controller.
- Centralised control and global view of the network that is achieved with multiple physical or virtual instances behaving like a single element.
- Open interfaces between devices in both planes to promote interoperability and flexibility to stimulate innovation.
- Programmability of the network as a single entity by external software controller and open interfaces.

An overview of the architecture is often mentioned in the literature covering four key interfaces and divided into three major layers. Figure 2.4 shows both of these aspects, based on [JZHT14] and [SSCF13].

Regarding layers, the bottom one is the Infrastructure Layer that involves both physical and virtual equipment as routers and switches (data plane). The Control Layer is composed of controllers responsible for managing the different flows and paths in the networks and connects with the remaining layers through the north- and southbound Application Programming Interfaces (APIs). The top layer consists of all the different possible applications that can explore the already mentioned SDN benefits.

Concerning interfaces, the four following APIs are represented:

- The Southbound API, responsible for one of the main objectives of SDN, separating the control and data planes. The ONF defined the most popular standard for this interface called OpenFlow [ONFo13].
- The Northbound API, allowing the exchange of information between the controller platform and the applications running on top of the network. The key is to provide an abstraction of the network in the top layer.
- The Eastbound API, in charge of the translation module between SDN and legacy technologies such as Multi-Protocol Label Switch (MPLS). It guarantees full compatibility between different network implementations.
- The Westbound API, accountable for the exchange of information between controllers of different network domains.

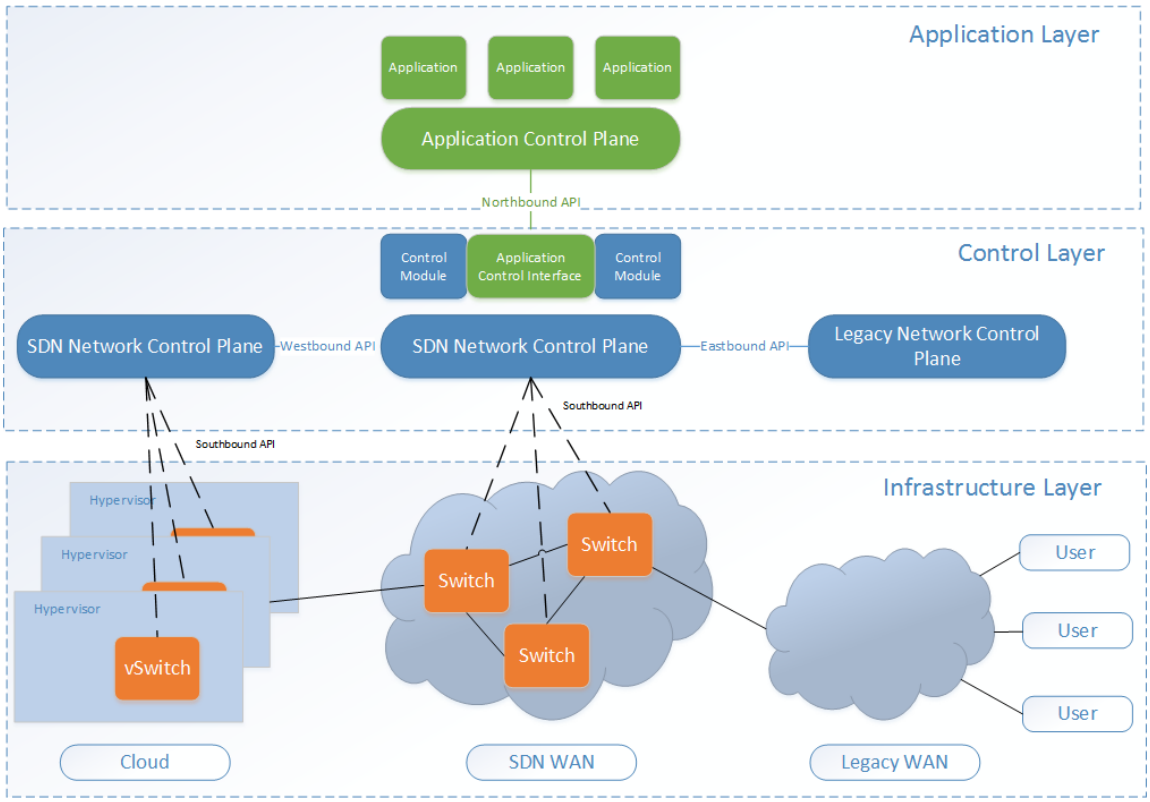

Figure 2.4. SDN architecture with layers and interfaces (adapted from [JZHT14]).

In [JZHT14], multiple uses of SDN are mentioned and described. Among others, use cases such as cloud orchestration, load balancing, routing and monitoring of the network are suggested. However, many challenges still need to be addressed. [SSCF13] refers to the performance and flexibility trade-off, security aspects and technical issues regarding interoperability and scalability as some of the current issues to be discussed.

2.2.2 Network Functions Virtualisation

This section addresses the concept of Network Functions Virtualisation (NFV). Benefits and challenges are discussed, as well as the distinction and complementarity between NFV and SDN.

In traditional networks, network operators are equipped with a large number of specialised hardware appliances, each one performing different functions for a single service. Besides that, service components have a strict ordering that is reflected in the network topology. Inability to adjust to rapid service changes due to innovation, high capital expenditure and operational expenditure and difficulty to accommodate the hardware in physical locations arise as the main problems that NFV aims to solve. To do that, NFV leverages standard IT virtualisation techniques to consolidate network equipment onto industry standard high volume servers, switches and storages, which could be located in data centres, distributed network nodes and end user premises [ETSI12]. Virtual Network Functions can then be implemented in software running on one or more physical servers achieving the decoupling between hardware and software applications. The European Telecommunications Standards Institute (ETSI) is supporting an operator-led Industry Specification Group (ISG) in order to develop NFV standards. Figure

2.5 shows a basic architecture of NFV applied to the EPC, an example of an application that has attracted attention from industry.

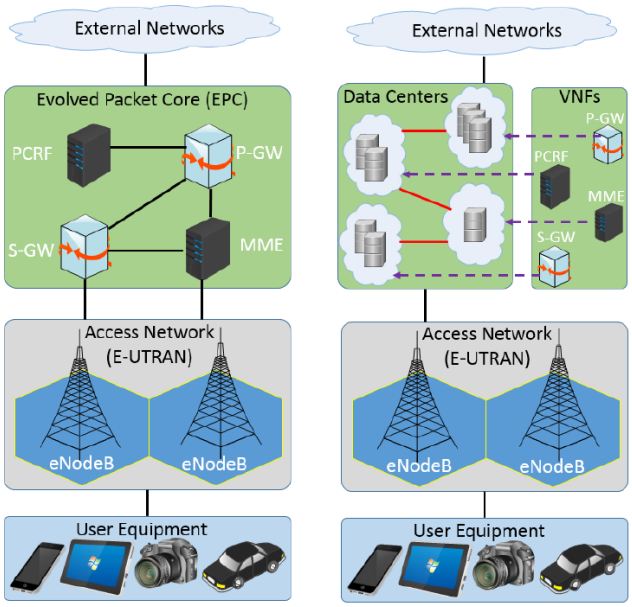

Figure 2.5. Virtualisation of the EPC (extracted from [MSGB15]).

Network Virtualisation (NV) is the concept that describes the group of technologies in which physical infrastructure resources are abstracted and sliced into virtual resources holding certain functionalities that are shared by several parties. NFV deploys the network functions over those virtual resources. Figure 2.6 represents a simple view of the NFV framework [ETSI13].

In [HSMA14] and [ETSI12], some benefits of NFV are pointed out, being the openness of platforms, scalability and flexibility, performance improvement, reduction of costs, increased testing efficiency, optimised real-time network configuration and multi-tenancy support (tailored services can co-exist). Multiple use cases and fields of applications of NFV are also mentioned in [ETSI12] and [HGJL15]. Several aspects around security, computing performance, VNFs interconnection, portability, legacy networks connection and management issues need to be discussed and some solutions and requirements must be stated [HSMA14].

SDN and NFV are distinct and non-dependent technologies, but are highly complementary and together can add great value to the networks. For instance, the functions of the SDN controller can run on a VM, meaning that the SDN applications are performed as VNFs, and hence take advantage of NFV features. In an opposite way, SDN can accelerate NFV deployments through its benefits, such as automated configurations, policy control and security functions. Cloud computing is also an important concept related to both of these technologies, being one of the bases of NFV and relying on improved efficiency provided by an SDN technology. Distributed Management Taskforce (DMTF) has standards and work groups addressing cloud computing. OpenStack is a known software to control computational pools. Figure 2.7 represents the relation between the described concepts.

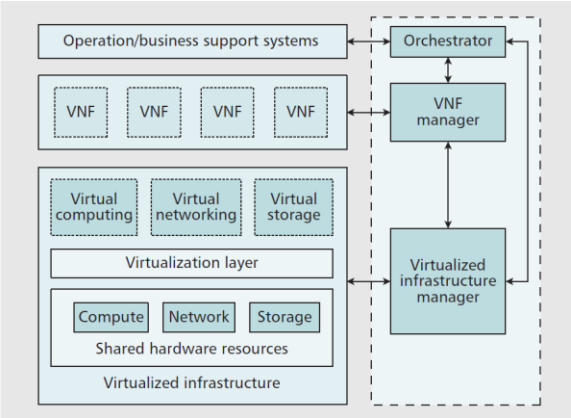

Figure 2.6. NFV architectural framework (extracted from [HGJL15]).

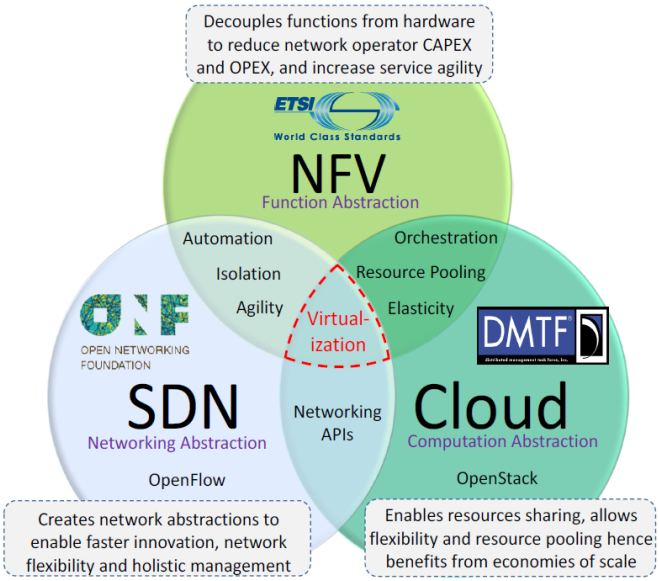

Figure 2.7. SDN, NFV and cloud computing relation (extracted from [MSGB15]).

2.2.3 Cloud Radio Access Network

This section addresses one of the main topics of this thesis. It provides a background on the key principles of Cloud Radio Access Network (C-RAN).

C-RAN is inspired by the current IT cloud computing solutions and consists of a novel radio access network where baseband resources are pooled so that they can be shared among base stations. The main reason of this approach is to design a less expensive and a more energy efficient solution for RANs, while offering the possibility to design a network with improved capacity based on resource sharing and adaptation to non-uniform traffic [CCYS14]. Besides Cloud RAN, the C-RAN concept also refers to the centralised processing, cooperative radio and clean system (often referred as Green RAN).

The idea emerged of the already proposed architecture, in which the base station is separated into a remote radio unit (RRH) and a baseband signal processing part (BBU), which is sometimes mentioned as Data Unit (DU). RRH is the component responsible for power amplification and analogue processing, and may do other functions, such as filtering, digital to analogue and analogue to digital conversions or

pre-distortion. These last functions depend on the technological splitting point chosen. BBU deals with the common digital baseband functions of a base station, such as modulation and coding. This solution already leads to lower power consumption and more convenient placement of BBUs. The innovation presented by C-RAN intends to optimise the BBU utilisation in base stations. To achieve that, a BBU pool (a cluster of baseband units) is virtualised and shared among cell sites. A C-RAN architecture applied to an LTE network is represented in Figure 2.8. The X2 interface is redefined (known as X2+) and offers inter-cluster communication.

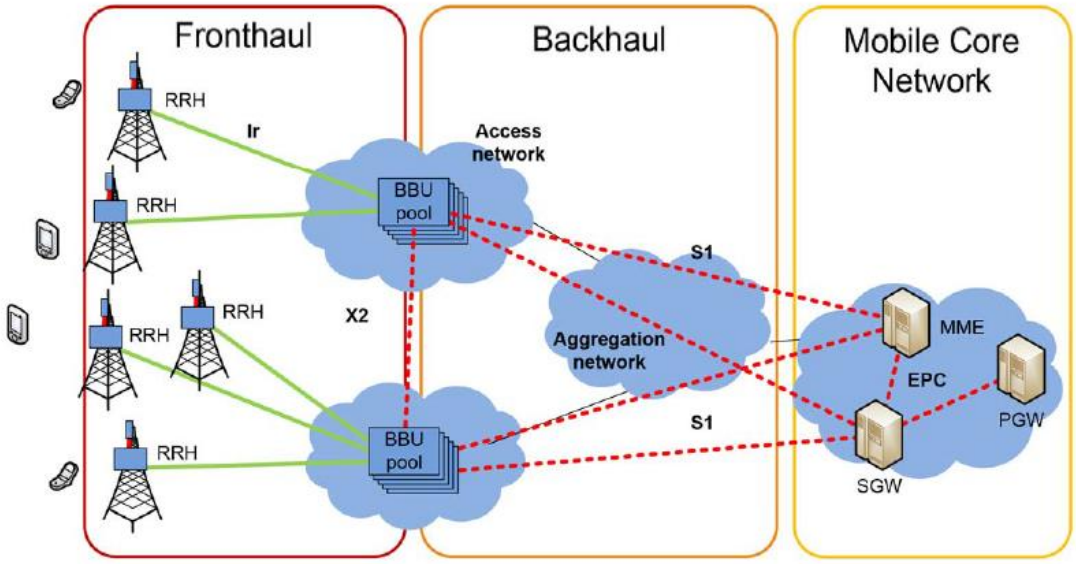

Figure 2.8. C-RAN in an LTE network (extracted from [CCYS14]).

Two implementation schemes of C-RAN based on two different splitting methods of functions are often mentioned in the literature. In [CMRI11], the two proposed solutions are the following:

- Fully Centralised Solution, where functionalities from layers 1 through 3 (L1, L2 and L3) are implemented in the BBU. This solution appears as an advantage, because it simplifies resource sharing in the BBU pool and facilitates the implementation of collaborative advanced functions, such as CoMP, allowing a significant improvement in network capacity. However, this solution requires high bandwidth connections to transport the digital I/Q data between RRH and BBU.
- Partially Centralised Solution, in which L1 functions are concentrated in the RRH. This approach introduces some complexity in the RRH, leading to a less flexible and upgradable implementation. Besides that, collaborative functionalities become harder to implement and the L2-L1 interaction is considered to be complex [PWLP15]. The great advantage of this solution is that unmodulated data can be 20 to 50 times of the original baseband I/Q data [CMRI11], lowering the transport network burden.

Other hybrid solutions are proposed. [NGMN13] presents five use cases for functions allocation.

Besides the BBU pool and the RRHs, the transport network is referred to as the third element of the C-RAN architecture, due to its global importance. In a cost saving concept like C-RAN, the transport

network is the part to be improved, being likely to be the technology in which operators will have to invest more money. High bandwidth, low latency and low jitter are required for the fronthaul in order to achieve LTE and LTE-A strict restrictions related to acknowledgment protocols. In order to attain switching capability, this transmission is done over the defined Common Public Radio Interface (CPRI).

The main advantages related to C-RAN are (based on [CCYS14]):

- Energy efficiency and cost savings. A centralised BBU pool reduces the power consumption mostly related to air-conditioning and processing equipment [CMRI11].
- In addition, interference reduction techniques as well as simpler RRH implementation (only the antennas and feeders are needed) allow for small cells deployment with less energy used in the BS and in the UE. Cost savings can also be achieved by reducing civil work and gathering all equipment in a central room. Besides that, resource sharing among BSs allows to adapt to traffic demands and reduce power without compromising the required system availability.
- Scalability and adaptability. Instead of designing each base station to the peak traffic load, a C-RAN technology allows for a better overall utilisation rate with resource sharing in the BBU pool (Statistical Multiplexing Gain). With this, large scale traffic adaptability is achieved and capacity and coverage are easily improved with the addition of new cells. Load balancing can also be enabled with advanced algorithms to increase the network capacity and efficiency.
- Throughput improvement. The co-location of multiple BBUs in a pool eases the implementation of advances cooperative techniques, reducing processing needs and communications delays compared to a traditional architecture. Those techniques are designed to reduce interference (enhanced ICIC in LTE-A) and to increase the user's throughput (CoMP). Some of the possible carrier aggregation techniques in LTE-A also benefit from this architecture.
- Upgradability and ease of maintenance. As all processing units are centralised in a single location, fault recovery through automatic reconfiguration is possible. Besides that, the use of a Software-Defined Radio (SDR) approach eases the adoption of new standards with software upgrades.
- Possibility to perform traffic offload. An optional edge service can be offered, for example, trough Content Distribution Networks (CDN) or caching [NGMN13], reducing backhaul traffic and cost, and offloading traffic from the core network, providing a better user experience.
- Coexistence of multiple standards. Multimode base stations can be easily deployed with universal RRH [HDDM13].

Behind the mentioned benefits, there are still pending issues to be addressed in order to fully reach C-RAN objectives. The first one is to implement fronthaul solutions capable of achieving the already discussed fix requirements, without compromising the cost of the overall solution (one of the premises of the technology). Compression of data, architectural changes, optical fibre networks (for instance, WDM, dark fibres and PON [Fuji14]) and microwave links (e.g., working in the E-band) are the resolutions discussed in the literature. Besides that, BBU pool cooperation and interconnection raises challenges, such as optimal clustering of BBU to optimise the statistical multiplexing gain (MG), and efficient eICIC and CoMP algorithms supported by a secure and resilient network of flexible RRH

interconnection in the pool. In addition, virtualisation technologies based on IT cloud computing over general purpose processors (GPP) must be adapted to work real-time on processing capability allocation as any telecommunication system demands. At last, [CMRI11] states that edge services to offload traffic are also a challenge to C-RAN implementation.

SDN, due to its centralised and global view of the system, stands as a suitable approach to control self-organising networks (SON) to dynamically correspond to network changes (self-configuration, self-optimisation and self-healing), to deploy interference schemes (such as CoMP) and to dynamically program the RRHs connected to the pool (self-defining fronthaul) [ASRa15]. Regarding NFV, most of the work that has been made focuses on the core network of LTE. However, implementation in C-RAN will introduce flexibility between software and hardware of different vendors, allowing for simpler upgrades. Both of these concepts are compatible and beneficial in a C-RAN context. Figure 2.9 exemplifies a C-RAN architecture in which SDN participates both as a controller of the transport networks and as a C-RAN controller. NFV is responsible for firewall and virtual EPC elements.

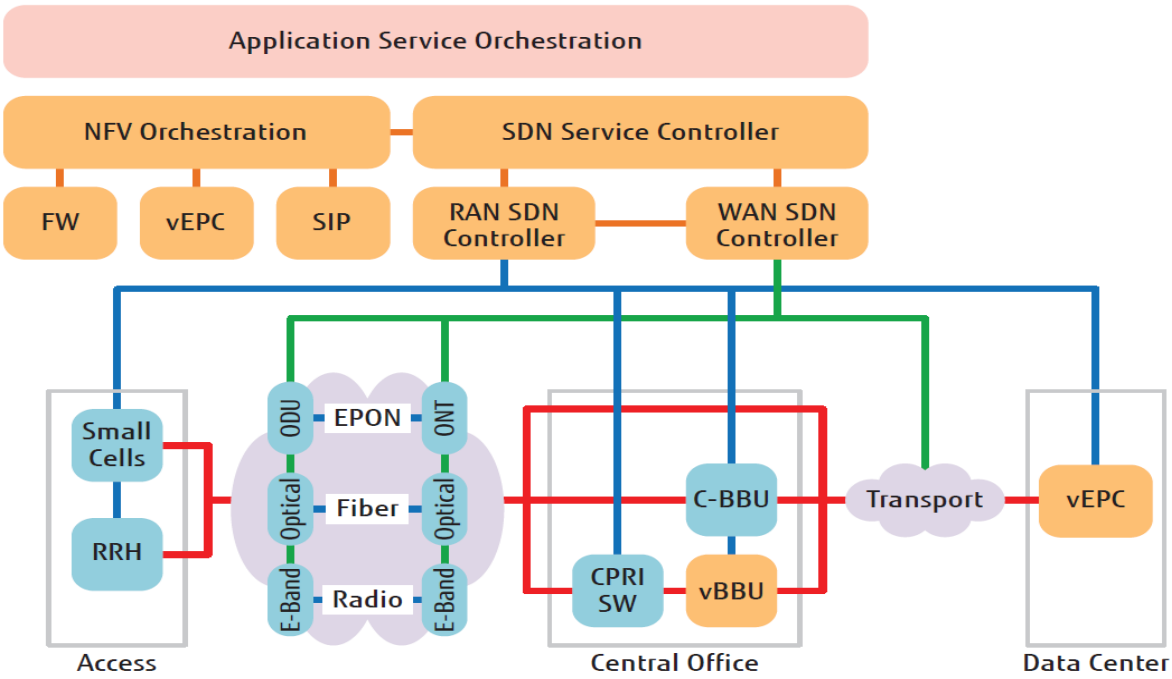

Figure 2.9. C-RAN architecture working with SDN and NFV (extracted from [Fuji14]).

2.3 Small Cells

This section describes the main concepts of small cells used in LTE-A systems and their integration in a possible C-RAN deployment.

Small cells enhancements are a new development trend that has been attracting operators' attention to handle the capacity requirements of high-traffic areas. Their deployment in indoor and outdoor scenarios (introduced in Release 12 of 3GPP [3GPP13a] [3GPP13b]) focuses on the densification of the network

through the installation of low-cost base stations to serve a smaller number of users in each one of them, resulting in a heterogeneous network.

The two main goals in using this technology are the possibility to handle higher traffic demands in specific areas (hotspots), and the capability to provide additional coverage in areas that are not filled in by the macro base stations. Instead of adding more sectors or more eNodeBs in the overall planning, operators can deploy low-power base stations or RRH with reduced costs in site acquisition and equipment. With this, traffic is offloaded from the macro-cell network and the overall network performance grows with higher bit rates per user achievable and improved coverage zones.

This kind of system can be implemented with or without macro-cell coverage. In the first scenario, the bandwidth used in the smaller cells can be different, or it can be the same as the one used in the macro-cell base station. In the latter situation, the utilisation of the radio resources is maximised with a frequency reuse of 1. Besides that, small cells can be deployed sparsely to cover hotspots or densely to improve the throughput of a larger zone. The most common implementations of small cells are pico-cells, Home eNodeBs (femto-cells), relay nodes and trusted wireless local area networks, such as WiFi. Small cells can also be implemented through RRHs in a centralised RAN architecture.

Besides the described advantages, the increased proximity among cells in this architecture imposes technical requirements, such as interference coordination among small cells and between small and macro-cells, joint processing techniques, robust mobility for frequent handovers and a low latency backhaul to handle higher throughputs [NNBK13].

As previously mentioned regarding the C-RAN implementations in LTE-A, metropolitan areas with a diameter of up to 40 km (due to latency restrictions [CCYS14]) are also seen as the most advantageous implementation in terms of statistical multiplexing gain.

From the main advantages described in Section 2.2.3 for C-RAN systems, small cells requirements can benefit from the described architecture. In this case, the signalling needed between small and macro-cells to control interference or to add improved cooperative signal transmission can be performed more efficiently through BBU pools. In addition, this kind of solution imposes fewer restrictions in the shorter transport needs and simplifies the physical implementation with even smaller and quicker physical deployment, in most of the cases without giving up of the legacy rooms. Multiple scenarios are often described as typical situations in which small cells appear as an advantage [CCYS14]:

- Heterogeneous Networks (HetNets) where small cells are deployed with new low-power RRHs.
- Overlay networks to boost system capacity using RRHs working in a different frequency band than the macro cell.
- Super-hotspots (e.g., stadiums or transportation hubs) covering highly populated areas to ensure capacity and coverage to mobile users.
- Small cells replacing a macro-cell to increase system capacity and to improve and simplify the overall coverage (e.g., indoor coverage with one RRH per floor [NGMN13]).
- Highways and railways in which frequent handovers can be performed efficiently within the BBU pool.

The mentioned scenarios are represented in Figure 2.10: the multiple fronthaul connections are exemplified, as well as the access ring using an Optical Transport Network (OTN) or Carrier Ethernet. Another concept suggested in this figure is RAN as a Service (RANaaS), in which multiple operators share the RRH infrastructure, the fronthaul connections and possibly the BBU pool computational resources. Actually, this kind of business model can be easily implemented with the aforementioned splits in the architecture.

All previous cases can be combined to provide different solutions and all of them should rely on techniques such as eICIC and CoMP to assure the proper conditions in a denser cellular environment.

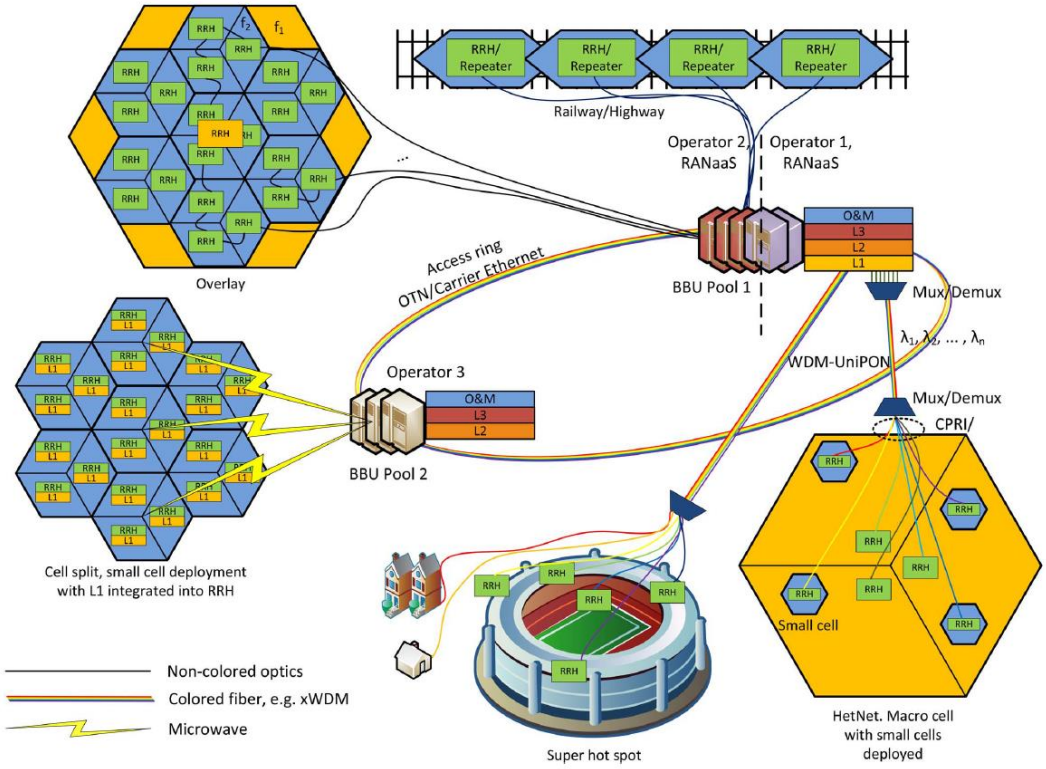

Figure 2.10. C-RAN deployment scenarios (adapted from [CCYS14]).

2.4 State of the Art

Several studies addressing the major challenges imposed to C-RAN deployments in LTE-A systems (mentioned in Section 2.2.3) are described in the literature. This section states the work developed by several authors in the area of implementation analysis and performance of C-RAN architectures.

In [ASBD15], the critical timing issues of C-RAN in LTE FDD are identified. The authors use a total delay budget of 3 ms at the eNodeB due to the acknowledgment message mechanism and consider a maximum latency of 400 µs for the digital fronthaul transmission over fibre. For the remaining timing budget, the authors propose a software and hardware setup configuration of virtual machines running

on GPPs and reach a result of 99.5% of subframes processed within their proposed deadline. The required processing power is also estimated in this work.

Regarding the statistical multiplexing gain achieved with the centralisation of processing resources into virtual base stations, the authors in [LZGN14] introduce a mathematical model to fully analyse it. In this work, a method to calculate the blocking probability of a virtual base station pool and the limit of the pooling gain are derived. The results confirm the statistical multiplexing gain and show that larger pools achieve a negligible gain that might not justify larger economical investments. The model does not cover features such as CoMP, and the authors state that verification with realistic data is needed.

Working on the statistical multiplexing gain as well, the authors in [ChHC14] analyse the traffic profiles of office and residential areas, in order to evaluate the pooling strategies that lead to higher pooling gains and higher potential cost savings. The results point to an achievable multiplexing gain up to 1.6. In this scenario, 20%-30% of the office cells are pooled together with 70%-80% of residential cells. In the same work, the authors also compare the cost of one BBU and one kilometre of fibre to reach the conclusion that the most advantageous deployment situations are the dense urban scenarios (less than 100 km^2). In [ChCB13], a similar approach is taken and the simulation results in the city of Cologne with traffic prediction for 2017 prove a 75% reduction of BBU resources compared to traditional RAN architectures. The authors state that the value is higher than expected and that 50% of reduction should be the most optimistic scenario in real situations. The same authors propose in [HCAC15] a linear programming method to show how to optimally assign cells to multiple BBU pools when maximising the statistical gain and minimising the fibre length needed. In addition, the authors in [WeGP13] study the multiplexing gain achieved with processing capacity instead of data rates or traffic, using a power model to characterise the RRHs capacity in Giga Operations per Second (GOPS). The results obtained point to savings of 15% aggregating 57 sectors in the same pool, and show a strong dependence on the user distribution throughout the scenario. Using the same power model, [Maro15] builds a model to analyse the overall capacity in GOPS in the city of Lisbon, although does not add traffic variation during the day. The same author builds a proliferation model to analyse the influence of small cells. Still, the authors in [BCJK12] build a framework for experimental results on processing savings resulting in about 27% of savings, but concluding that only 2% are derived of the centralisation. The main cause of savings is the under-provisioning of resources that could be also done in a non-centralised implementation.

Based on the previously presented works, the authors in [AvCI15] apply their model to different case studies including heterogeneous network deployments and different traffic profiles. An optimal ratio of 22% of office cells pooled with 78% of residential cells is obtained. The analysis is done based on cost and sensitivity to traffic variations. The authors also point out the benefits of introducing a dynamic BBU-RRH mapping to achieve higher gains and state that their model could be adopted for the re-assignment strategies. NFV and SDN solutions are suggested for the dynamic fronthaul.

In [NaWK12], a semi-static and adaptive BBU-RRH switching scheme are proposed in order to add flexibility in the assignment of BBU resources to accommodate the peak hour traffic. The results prove that the number of BBUs is reduced 26% for the semi-static approach and 47% for the adaptive one. In [LSJR13], a flexible fronthaul is also proposed with the goal to improve the overall performance and

energy efficiency. With a WiMAX experiment with small cells, the results relative to the users' throughput confirm the need for a reconfigurable fronthaul. Radio-over-fibre (RoF) is the transmission technology used in the testbed.

In [SASR15], a fronthaul framework entitled *FluidNet* is also proposed to capture the different traffic patterns of small cells. The results show a 50% improvement in satisfying the traffic demands and 50% of reduced computational resources in the BBU pool compared to the baseline schemes. A WiMAX C-RAN testbed was implemented to prove the concept of the proposed framework. The fronthaul transmission used was radio-over-fibre instead of a digital transmission method.

The spectral efficiency is also an issue described in the literature. In [CJLS14], a simple cooperative transmission scheme among RRHs in a HetNet environment achieves significant throughput improvements compared to a local implementation. In the same work, a power model is used to compare the energy efficiency of a C-RAN implementation to a regular HetNet scenario based on the activity of the users in the residential and office areas. The simulation model shows promising results related to energy efficiency when the users' activity is mostly generated in the residential cells due to lower interfering scenarios. Also related with HetNet, the authors in [SuRF15] propose a complex model to study the cost effectiveness of heterogeneous C-RAN networks. The study includes fronthaul costs and savings of 15% per square kilometre are achievable. A mixture of fibre and microwave links is also suggested to reduce the costs of C-RAN implementation.

In [LLXB12], a field test in Beijing is described. In this field testbed, the overall throughput is evaluated for the UL with and without a cooperative multipoint reception. The obtained results prove that there are significant throughput gains when the user is at the cell edge. The CoMP mechanisms used are diversity combiners, namely the Maximum Ratio Combiner and the Interference Rejection Combiner. The latter combiner registers better results in the field tests with interference from another user.

Regarding the fronthaul connection, in [GLDP12] a microwave link working in the E-band (70-80 GHz) and a wavelength division multiplexing-passive optical network (WDM-PON) were tested in Beijing to transport traffic from a 2.5 Gbps CPRI link. This testbed was conducted to proof the concept of heterogeneous networks built as C-RAN. A theoretical analysis can also be found in [MBCT15], where the authors studied the overlay and OTN options for fronthaul, optimising the switching and BBU Pools locations in a mixed combination of fronthaul and backhaul links.

In [ACHC15], an analysis considering only CAPEX is done to estimate the feasibility of using cheaper microwave links to replace optical fibres. This study considers a cost factor depending on the population density and the average traffic per user. The results show that for high density areas the microwave replacement leads to lower CAPEX. However, the single cell peak data rate is limited due to microwave capacity restriction. The authors also suggest a solution in which eNodeBs and RRH are simultaneously connected to the BBU pool. This last concept simplifies the fronthaul/backhaul connection but it reduces the achievable multiplexing gain.

Chapter 3

Models and Simulator Description

This chapter contains the description of the algorithms used to develop the model and the parameters to be analysed. Section 3.1 presents a brief description of the problem and the structure of the model. Section 3.2 defines the parameters to be studied. Section 3.3 presents the algorithms to be used. Finally, Section 3.4 does the model assessment to ensure the relevance of the model.

3.1 Model Description

3.1.1 Problem Definition

The problem addressed in this thesis concerns the assignment of RRHs to BBU pools in an urban scenario. The goal is to study multiple performance parameters of a C-RAN implementation in the current deployed LTE-A network.

As previously mentioned, the new transport segment known as fronthaul must introduce a limited delay in order to comply with the most stringent requirements of LTE-A. This limitation imposes a maximum length between the links connecting RRHs to the pools. With that being said, it is important to measure and quantify the delay of the links in the scenario under study.

Another important goal is to study the load balancing between the possible pools' locations. It is essential to understand how the processing power is being assigned over the scenario, and to conclude about the implications of this metric in network dimensioning.

To quantify the already discussed multiplexing gain, one should also measure the achievable gains and hardware savings. It is important to establish relations of this value with the scenario under study, namely with the number of office and residential cells deployed. The results have an impact on the hardware savings achieved with centralisation. Finally, one should study the option to maximise centralisation, by using the minimum number of pools to serve all RRHs.

Besides all technical parameters under study, it is also important to study and to quantify how cost efficient is C-RAN technology. The concept presents theoretical gains in both CAPEX and OPEX, these gains requiring to be measured.

As described in Section 2.3, C-RAN can also be associated with multiple small cell deployment scenarios. Inspired by this idea, a proliferation algorithm was also implemented in this thesis, to forecast the growth of RRHs and of traffic demand in future years. Naturally, a proliferation study has an impact on BBU requirements, on the pooling of different traffic profiles, and on links capacity.

3.1.2 Model Structure

In order to solve the problem mentioned in Section 3.1.1, one has structured a simulator with three main components represented in Figure 3.1:

- Proliferation layer.
- Technical layer.
- Cost layer.

The first component is the **proliferation layer**. This phase of implementation is developed to add a temporal dimension to the simulator. The technology under discussion being a futuristic concept, one

uses the proliferation algorithm to study how the architecture can scale and adapt to future network demands. Taking the current RRHs positions and the number of years to proliferate as input, this layer adds RRHs coordinates based on an adjustable proliferation factor.

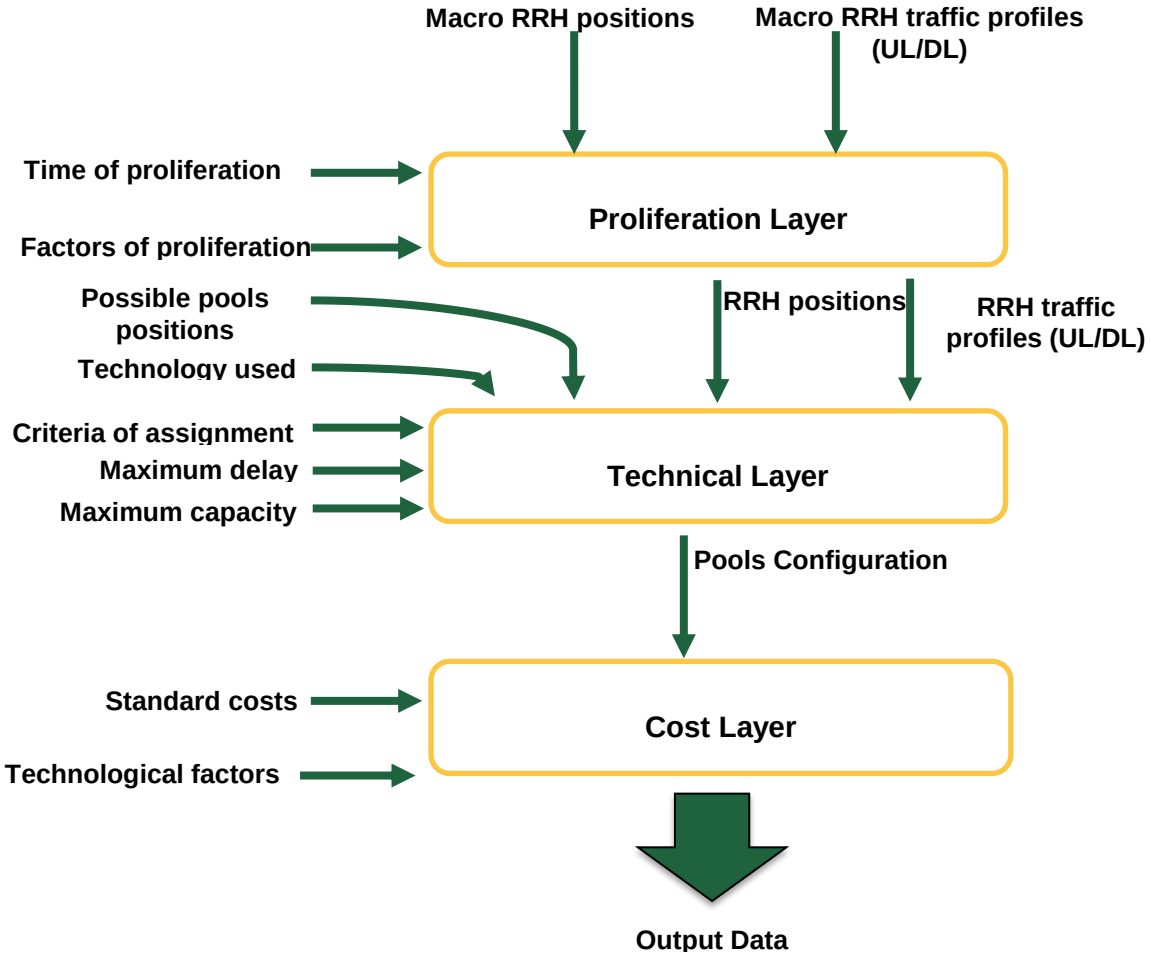

Figure 3.1. Model structure.

The second component is the **technical layer**, which was developed to make RRH-BBU pool assignments based on different metrics. In this phase of implementation, one can select one of two technological implementations. One is the non-virtualisation scenario that reflects the current state of the art of the physical equipment in the market, in which case a centralised implementation is useful for cooperative transmission techniques and for eventual pooling gains in which some cells can share a portion of the hardware resources. Note that this pooling gain is not the same as the virtualisation gain, as just some of the physical functions are shared, and only between some sub-clusters of cells; besides that, the centralised implementation presents dimensioning gains even without virtualisation, and in the mentioned scenario, one can use metrics such as minimise delay and balance the load (number of RRHs per pool). The other is the virtualisation one that projects the future implementation of C-RAN in which all the processing functions can be virtualised and implemented in general purpose processors; in the mentioned scenario, one can use the same metrics as in the non-virtualisation approach, as well as new ones taking the temporal variation of the traffic demand or processing power among all the RRHs into account. In the latter case, one uses an assignment strategy to maximise virtualisation and

multiplexing gain.

The third and final component of the model is the **cost layer.** In this phase, one develops a model to compare costs in a traditional implementation and in a centralised one. Comparison results show possible savings obtained with the adoption of C-RAN technology instead of local RAN (green field deployments). This layer also offers different factors to account for possible computational implementations of the data centres.

3.2 Performance Parameters

3.2.1 Latency

Latency or delay is an important metric in telecommunication systems. Any C-RAN implementation must be compliant with latency limits standardised for the data plane, control plane and synchronisation in order to guarantee the correct implementation of LTE functionalities. As the main difference introduced by centralised architectures is the split of the base station into RRH and BBU with the addition of fronthaul connections, the delay limit to be achieved corresponds to the maximum latency allowed for processing in traditional eNBs. In C-RAN, the distributed elements of the base station introduce delays as it is expressed in the following expression:

$$\delta_{eNB\,[\mathrm{ms}]} = \delta_{RRH\,[\mathrm{ms}]} + \delta_{RTT\,[\mathrm{ms}]} + \delta_{BBUpool\,[\mathrm{ms}]} \tag{3.1}$$

where:

- δ_{eNB} – Maximum delay budget for eNB.
- δ_{RRH} – Delay budget for RRH processing.
- δ_{RTT} – Round Trip Time for fronthaul transmission (two-way delay).
- $\delta_{BBUpool}$ – BBU pool processing delay.

The eNB latency can be derived from 3GPP specifications [3GPP14]. The timing of the data plane Hybrid Automatic Repeat Request (HARQ) retransmission process is the critical one in FDD-LTE. The overall process is represented in Figure 3.2, where Transmission Time Interval (TTI), frame alignment and processing time in UE and eNB are presented. The total delay experienced by the end user depends on the probability of retransmissions due to transmission errors as well.

The processing budget of a traditional eNB used to estimate the overall delay is defined as follows, corresponding to the sum of UL and DL processing times,

$$\delta_{eNB\,[\mathrm{ms}]} = 2.0\ \mathrm{ms} \tag{3.2}$$

The addition of the restriction in (3.2) to the delay budget of C-RAN components may be used to estimate the maximum distance achievable for a link. In (3.3), one shows the dependence of the maximum distance with propagation speed and tolerated delay.

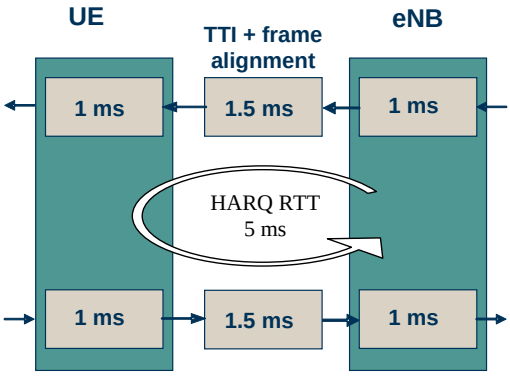

Figure 3.2. Timing restrictions for LTE-A (extracted from [3GPP14]).

$$d_{FH\,[\mathrm{km}]} = v_{[\mathrm{km/ms}]} \frac{\delta_{RTT\,[\mathrm{ms}]}}{2} \tag{3.3}$$

where:

- v – Transmission speed in the link.
- d_{FH} – Fronthaul allowed distance.

The achievable distances are medium dependent, as the transmission speed differs if the transport solutions chosen for the fronthaul is optical fibre or microwave links.

In [PCCR13], a total up to 400 µs is considered for the two-way delay of the fronthaul connection in LTE-A. However, a more stringent value of 200 µs is often considered in the literature to account for delay sensitive functions, such as CoMP.

3.2.2 Link Capacity

Fronthaul link capacity is the traffic generated by a cell site. The value measured for this parameter depends on the number of RRHs serving the cell site. It is important to estimate this value in order to characterise the links needed between RRHs and the BBU pool. As discussed in Chapter 2, the most widely referred protocol is CPRI. In [PCCR13], equation (3.4) is used to calculate the data rate per CPRI link corresponding to 1 carrier and 1 sector; note that an 8B/10B line code is used in the link and that the framing efficiency is also considered for the overall throughput.

$$R_{CPRI\,[\mathrm{Mbps}]} = N_{MIMO}\, f_{s\,[\mathrm{Msamples/s}]}\, N_{\mathrm{b}\,[\mathrm{bit/sample}]}\, 2\, \frac{1}{C_{line}}\, \frac{1}{\eta_{framing}} \tag{3.4}$$

where:

- R_{CPRI} – Data rate required in the CPRI link for the useful data;
- N_{MIMO} – Number of antennas per sector (MIMO);
- f_s – Sampling rate;
- N_b – Sample width;
- C_{line} – Line encoding used (8B/10B or 64B/66B);
- $\eta_{framing}$ – Framing efficiency (16/15);

The values presented in Table 3.1 are the most relevant for this thesis. The value for LTE 100 MHz 8x8 RRH is presented to explore the effect of carrier aggregation in LTE base stations and high-order MIMO configurations. However, this last value is not specified in the current version of CPRI standards.

Table 3.1. Typical data rates of CPRI links (adapted from [PCCR13]).

RAN configuration	LTE 10MHz 2x2	LTE 10MHz 4x4	LTE 20MHz 2x2	LTE 20MHz 4x4	LTE 100MHz 8x8
CPRI data rate [Gbps]	1.228	2.458	2.458	4.915	49.15

The values presented in Table 3.1 correspond to a current limitation in C-RAN deployments. Time-based compression schemes with compression factors of 2 to 3 can be implemented without introducing significant extra delay. However, higher compression factors are needed, and are being studied to ensure that CPRI links can handle functionalities such as 8x8 MIMO and carrier aggregation.

3.2.3 Processing Capacity

The need to model the power required by each base station comes from the virtualisation of base stations in the BBU pool. In order to fully analyse the implementation of C-RAN, the overall processing capacity required for a given network configuration should be considered. This parameter is measured in Giga Operations per Second (GOPS). In [DDLo15], a power model is used to characterise the processing capacity required for different types of LTE base stations. A simple adaptation of the proposed model to a centralised architecture such as C-RAN is described as:

$$P_{pool\,[\mathrm{GOPS}]} = \sum_{i=1}^{N_{BBU}} P_{BBU,n\,[\mathrm{GOPS}]} + P_{fixed\,[\mathrm{GOPS}]} \tag{3.5}$$

where:

- P_{pool} – BBU pool processing power.
- $P_{BBU,n}$ – Single BBU processing power.
- P_{fixed} – Fixed processing power for scheduling and signalling, independent of the number of BBUs.
- N_{BBU} – Number of BBUs in the pool in pool p.

The model proposed in [DDLo15] presents the tools needed to estimate the baseband processing power that can be associated with each BBU instance in a BBU pool for DL and UL. The model takes the physical layer processing and the communication protocols in the second layer of LTE into account. Note that one is assuming a splitting point in C-RAN, where all digital functions are centralised. The total power consumption is then given by the processing powers required in the physical layer, for data flows management and system control, for high-level protocols of layer 2, and for backhauling to the core network:

$$P_{BBU,n[\mathrm{GOPS}]} = P_{phy[\mathrm{GOPS}]} + P_{control[\mathrm{GOPS}]} + P_{HLP[\mathrm{GOPS}]} + P_{BH[\mathrm{GOPS}]} \tag{3.6}$$

where:

- P_{phy} – Processing power required for the physical layer functions.
- $P_{control}$ – Processing power required for management of data flows, scheduler and system.
- P_{HLP} – Processing power required for high-level protocols in LTE processing stack.
- P_{BH} – Processing power required for the S1 interface depending on the S1 data rate.

The physical processing power depends on multiple digital processing functions:

$$P_{phy[\text{GOPS}]} = O_{1[\text{GOPS}]}{P_f}^{DPD} + O_{2[\text{GOPS}]}{P_f}^{filter} + O_{3[\text{GOPS}]}{P_f}^{sampling} + O_{4[\text{GOPS}]}{P_f}^{TD} + O_{5[\text{GOPS}]}{P_f}^{FFT} + O_{6[\text{GOPS}]}{P_f}^{MIMO} + O_{7[\text{GOPS}]}{P_f}^{sync} + O_{8[\text{GOPS}]}{P_f}^{channel} + O_{9[\text{GOPS}]}{P_f}^{equaliser} + O_{10[\text{GOPS}]}{P_f}^{equalisation} + O_{11[\text{GOPS}]}{P_f}^{OFDM}\ O_{12[\text{GOPS}]}{P_f}^{Mapp} + O_{13[\text{GOPS}]}{P_f}^{coding} \quad (3.7)$$

where:

- ${P_f}^{DPD}$ – Digital Pre-Distortion scaling factor.
- ${P_f}^{filter}$ –Filtering scaling factor.
- ${P_f}^{sampling}$ – Up/Down scaling factor.
- ${P_f}^{TD}$ - Time domain functions for estimation and compensation scaling factor.
- ${P_f}^{FFT}$ - Frequency domain functions for FFT and IFFT scaling factor.
- ${P_f}^{MIMO}$ – MIMO precoding scaling factor.
- ${P_f}^{sync}$ – Synchronisation functions scaling factor.
- ${P_f}^{channel}$ – Channel estimation and interpretation scaling factor.
- ${P_f}^{equaliser}$ – Equaliser compensation scaling factor.
- ${P_f}^{equalisation}$ – Equalisation scaling factor.
- ${P_f}^{OFDM}$– OFDM modulation and demodulation specific functions scaling factor.
- ${P_f}^{Mapp}$ – Mapping and demapping functions scaling factor.
- ${P_f}^{coding}$ – Forward Error Correction functions scaling factor.
- O_x – Complexity associated with each function measured in GOPS, based on the reference scenario (see Annex A).

The scaling factors in (3.7) are computed through the reference scenario as follows:

$${P_f}^{DPD} = {P_f}^{filter} = {P_f}^{sampling} = {P_f}^{TD} = \frac{B_{[\text{MHz}]}}{B_{ref[\text{MHz}]}}\frac{N_A}{N_{A,ref}}\left(\frac{Q_{[\text{bits}]}}{Q_{ref[\text{bits}]}}\right)^{1.2} \quad (3.8)$$

$${P_f}^{FFT} = \left(\frac{B_{[\text{MHz}]}}{B_{ref[\text{MHz}]}}\right)^{1.2}\frac{N_A}{N_{A,ref}}\left(\frac{Q_{[\text{bits}]}}{Q_{ref[\text{bits}]}}\right)^{1.2} \quad (3.9)$$

$${P_f}^{MIMO} = \frac{B_{[\text{MHz}]}}{B_{ref[\text{MHz}]}}\frac{N_A}{N_{A,ref}}\frac{F_{DC[\%]}}{F_{DC,ref[\%]}}\frac{N_{streams}}{N_{streams,ref}}\left(\frac{Q_{[\text{bits}]}}{Q_{ref[\text{bits}]}}\right)^{1.2} \quad (3.10)$$

$${P_f}^{sync} = \frac{N_A}{N_{A,ref}}\left(\frac{Q_{[\text{bits}]}}{Q_{ref[\text{bits}]}}\right)^{1.2} \quad (3.11)$$

$${P_f}^{channel} = \frac{B_{[\text{MHz}]}}{B_{ref[\text{MHz}]}}\frac{N_A}{N_{A,ref}}\left(\frac{F_{DC[\%]}}{F_{DC,ref[\%]}}\right)^{0.5}\frac{N_{streams}}{N_{streams,ref}}\left(\frac{Q_{[\text{bits}]}}{Q_{ref[\text{bits}]}}\right)^{1.2} \quad (3.12)$$

$$P_f^{equaliser} = \frac{B_{[\mathrm{MHz}]}}{B_{ref[\mathrm{MHz}]}} \left(\frac{N_A}{N_{A,ref}}\right)^3 \frac{F_{DC[\%]}}{F_{DC,ref[\%]}} \left(\frac{Q_{[\mathrm{bits}]}}{Q_{ref[\mathrm{bits}]}}\right)^{1.2} \quad (3.13)$$

$$P_f^{equalisation} = \frac{B_{[\mathrm{MHz}]}}{B_{ref[\mathrm{MHz}]}} \left(\frac{N_A}{N_{A,ref}}\right)^2 \frac{F_{DC[\%]}}{F_{DC,ref[\%]}} \left(\frac{Q_{[\mathrm{bits}]}}{Q_{ref[\mathrm{bits}]}}\right)^{1.2} \quad (3.14)$$

$$P_f^{OFDM} = \frac{B_{[\mathrm{MHz}]}}{B_{ref[\mathrm{MHz}]}} \frac{N_A}{N_{A,ref}} \left(\frac{F_{DC[\%]}}{F_{DC,ref[\%]}}\right)^{0.5} \left(\frac{Q_{[\mathrm{bits}]}}{Q_{ref[\mathrm{bits}]}}\right)^{1.2} \quad (3.15)$$

$$P_f^{mapp} = \frac{B_{[\mathrm{MHz}]}}{B_{ref[\mathrm{MHz}]}} \left(\frac{E_{[\mathrm{bps/Hz}]}}{E_{ref[\mathrm{bps/Hz}]}}\right)^{1.5} \frac{F_{DC[\%]}}{F_{DC,ref[\%]}} \frac{N_{streams}}{N_{streams,ref}} \left(\frac{Q_{[\mathrm{bits}]}}{Q_{ref[\mathrm{bits}]}}\right)^{1.2} \quad (3.16)$$

$$P_f^{coding} = \frac{B_{[\mathrm{MHz}]}}{B_{ref[\mathrm{MHz}]}} \frac{E_{[\mathrm{bps/Hz}]}}{E_{ref[\mathrm{bps/Hz}]}} \frac{F_{DC[\%]}}{F_{DC,ref[\%]}} \frac{N_{streams}}{N_{streams,ref}} \left(\frac{Q_{[\mathrm{bits}]}}{Q_{ref[\mathrm{bits}]}}\right)^{1.2} \quad (3.17)$$

where:

- B – Bandwidth used in the BS.
- B_{ref} – Reference bandwidth used in the BS.
- N_A – Number of antennas in the BS.
- $N_{A,ref}$ – Reference number of antennas in the BS.
- $N_{streams}$ – Number of transmission streams.
- $N_{streams,ref}$ –Reference number of transmission streams.
- E – Spectral efficiency dependent on modulation and coding rate used.
- E_{ref} – Reference spectral efficiency dependent on modulation and coding rate used.
- F_{DC} – Frequency-domain duty cycling percentage (load).
- $F_{DC,ref}$ – Reference frequency-domain duty cycling percentage (load).
- Q – Number of bits used in quantisation.
- Q_{ref} –Reference number of bits used in quantisation.

The frequency-domain duty cycling is the fractional load of the system RBs, being used to quantify the effect of the load on the overall processing power. It is also possible to adjust the system load with a time domain duty cycle factor to represent the fraction of time in which the BS is sleeping for power savings. This last factor is actually important in modern base stations with power saving features.

The control and network processing powers are also described in [DDLo15], the backhaul power model not being derived in this work. One assumes:

$$P_{control} = \left(\frac{N_A}{N_{A,ref}}\right)^{0.5} \left(\frac{N_{streams}}{N_{streams,ref}}\right)^{0.2} \left(\frac{Q_{[\mathrm{bits}]}}{Q_{ref[\mathrm{bits}]}}\right)^{0.2} \quad (3.18)$$

$$P_{HLP} = \frac{B_{[\mathrm{MHz}]}}{B_{ref[\mathrm{MHz}]}} \frac{E_{[\mathrm{bps/Hz}]}}{E_{ref[\mathrm{bps/Hz}]}} \frac{F_{DC[\%]}}{F_{DC,ref[\%]}} \quad (3.19)$$

The presented equations are used with UL and DL reference tables in Annex A to calculate the number of operations per second required for each BS.

3.2.4 Multiplexing Gain

Multiplexing gain is a metric of the gains achievable with the centralisation of different traffic profiles. It is an important metric of one of the theoretical advantages of C-RAN, as explained in Section 2.2.3. Based on [ChHC14], one uses the following expression to quantify this parameter:

$$G_{mux,p} = \frac{\sum_{i=1}^{N_{RRH,p}} T^{peak}{}_{RRH,i[\text{GBph}]}}{T^{peak}{}_{pool,p[\text{GBph}]}} \tag{3.20}$$

where:

- $G_{mux,p}$ – Multiplexing Gain of pool p.
- $N_{RRH,p}$ – Number of RRHs connected to the pth pool.
- $T^{peak}{}_{RRH,i}$ – Peak traffic generated in the ith RRH.
- $T^{peak}{}_{pool,p}$ – Peak traffic handled by the pth pool.

When the goal is to quantify the overall multiplexing gain of multiple pools, the expression is:

$$G_{mux} = \frac{\sum_{i=1}^{N_{RRH}} T^{peak}{}_{RRH,i[\text{GBph}]}}{\sum_{p=1}^{N_{pools}} T^{peak}{}_{pool,p[\text{GBph}]}} \tag{3.21}$$

As for traffic per hour capacity, the gain can also be characterised by processing capacity. In this case, the gain is called processing gain, being computed by:

$$G_{proc} = \frac{\sum_{i=1}^{N_{RRH,p}} T^{peak}{}_{RRH,i[\text{GOPS}]}}{\sum_{p=1}^{N_{pools}} T^{peak}{}_{pool,p[\text{GOPS}]}} \tag{3.22}$$

A final gain can be defined that expresses not only the computational gain but also eventual capacity savings in terms of under-provisioning of cells, measuring the centralisation (as processing gain) and also the computational resources that are being wasted in decentralised implementations and that can be discarded in C-RAN with the scalability offered by data centres. The name of this parameter is the total gain, and it can be computed by:

$$G_{total} = \frac{N_{RRH} T^{peak}{}_{RRH[\text{GOPS}]}}{\sum_{p=1}^{N_{pools}} T^{peak}{}_{pool,p[\text{GOPS}]}} \tag{3.23}$$

where:

- $T^{peak}{}_{RRH}$ – Peak processing power required for an RRH under the most demanding radio conditions.

3.2.5 Dimensioning and Costs

One of the benefits often associated with C-RAN is its cost saving potential. These savings can be either in CAPEX or OPEX, as explained in Section 2.2.3:

$$C_{T[€]} = C_{CAPEX[€]} + yC_{OPEX[€]} \tag{3.24}$$

where:

- C_T – Total cost of the network.
- C_{CAPEX} – Investment costs.
- y – Number of years accounted for OPEX.
- C_{OPEX} – Operational expenses in one year.

As for investment expenses, one needs to account for the hardware cost, the licences paid and the cost of civil work required. One can also include the fronthaul costs as an initial investment:

$$C_{CAPEX[€]} = C_{HW[€]} + C_{CW[€]} + C_{lic[€]} + C_{FH[€]} \quad (3.25)$$

where:

- C_{HW} – Hardware cost.
- C_{CW} – Civil Work cost.
- C_{lic} – Expenses in licences.
- C_{FH} – Investment in fronthaul.

In [AHGr15], a hardware analysis for centralised RAN is evaluated. The authors of the work propose 3 types of technological implementations to be compared with a traditional approach in which the BBU is placed in a site and usually serving three cells (three RRHs).

The first implementation introduces the concept of centralisation. Instead of having a pool per site, one uses specific locations to accommodate for all processing functions. In this case, one assumes that there is no share of resources. The advantages of this deployment is that multiple BBUs can be stacked in the same rack while using only one power supply unit to serve them all, energy efficiency being expected to be higher. The second implementation introduces the concept of pooling, where some parts of the processing modules can be shared by a specific number of cells. Some switching schemes will certainly be needed to be accommodated. Finally, the third implementation assumes a full virtualised and C-RAN implementation, where the required processing cards are fully dependent on the gain introduced by the traffic variations in the scenario under study. A controller module is needed in comparison to previous solutions for dealing with all the complexity associated with virtual base stations.

As there is no relevant difference between the local implementation and the stacking one, only the former is considered, because the values can be estimated. The expression considers multiple hardware parts such as interfaces (backhaul and fronthaul), processing boards, physical infrastructure and energy equipment. For local implementation, the prices are normalised per cabinet serving a full cell site:

$$C_{HW,local[€]} = N_{cabinet}(C_{IO,BH[€]} + N_{FH/cab}\, C_{IO,FH[€]} + N_{BBboard} C_{BBproc[€]} + N_{Gboard} C_{Gproc[€]} + C_{control[€]} + C_{alarm[€]} + C_{Psup[€]} + C_{fan[€]} + C_{structure[€]}) \quad (3.26)$$

where:

- $C_{HW,local}$ – Normalised total cost of all the local BBU pools.
- $N_{cabinet}$ – Number of cabinets in the scenario.
- $C_{IO,BH}$ – Normalised cost of backhaul interfaces.
- $C_{IO,FH}$ – Normalised cost of fronthaul interfaces.

- $N_{FH/cab}$ – Number of fronthaul interfaces per cabinet.
- $N_{BBboards}$ – Number of boards used for processing in the local pool
- $N_{Gboards}$ – Number of boards used for processing in the local pool.
- C_{BBproc} – Normalised cost of baseband processing card.
- C_{Gproc} – Normalised cost of general processing card.
- $C_{control}$ – Normalised cost of a control unit.
- C_{alarm} – Normalised cost of an alarm unit.
- C_{Psup}– Normalised cost of a full power unit (supply and conversion).
- C_{fan} – Normalised cost of a fan unit.
- $C_{structure}$ – Normalised cost of the physical structure.

The pooling case introduces mostly cooperative transmission functionalities to the RRHs. As this thesis is not oriented to the radio component, it is not relevant to analyse this implementation. In this case, the same hardware components are used in the model. Some extra equipment related to switching and controlling of multiple BBUs is also taken into account. Besides, some factors are added to the model to study some extra complexity or price variations with future technology that is expected for C-RAN. The overall cost is computed by adding all the components required by rack in each pool.

$$C_{HW,CRAN} = \sum_{p=1}^{N_{Pools}} \Big(k_p \left(N_{BH/r} C_{IO,BH[€]} + N_{FH/r} C_{IO,FH[€]}\right) + \rho k_p V_p \big(N_{BBboards} C_{BBproc[€]} + N_{Gboards} C_{Gproc[€]}\big) + \nu N_{sh} \big(C_{control[€]} + C_{alarm[€]} + C_{Pconv[€]} + C_{fan[€]}\big) + \beta k_p C_{SW,r[€]} + \omega k_p C_{SW,l[€]} + k_p C_{controller[€]} + \theta k_p C_{Psupplier[€]} + k_p C_{rack[€]}\Big) \quad (3.27)$$

where:

- $C_{HW,CRAN}$ – Normalised total cost of one centralised BBU pool.
- $N_{BH/r}$ – Number of backhaul interfaces.
- $N_{FH/r}$ – Number of fronthaul interfaces per rack.
- k_p – Number of racks required in the pth pool.
- $C_{SW,r}$ – Normalised cost of one raw data switch.
- $C_{SW,l}$ – Normalised cost of one low-latency data switch.
- ρ – Cost reduction factor of GPPs.
- V_p – Virtualisation gain factor in the pth pool.
- β – Added complexity factor for raw latency switch.
- ω – Added complexity factor for low latency switch.
- ν – Added complexity factor for baseline components.
- N_{sh} –Number of shelves where baseline components are installed.
- θ – Added complexity factor for power units.
- C_{Pconv}– Normalised cost of a power unit (conversion).
- $C_{Psupply}$ – Normalised cost of a power unit (supplier).
- C_{rack} – Normalised cost of the physical structure.

Although this hardware model offers a complete dimensioning of the equipment, most of the required parameters are hard to estimate for a technology that is not mature yet. Besides that, the currently available equipment is usually sold as a whole, the price of individual components not being known. With that being said, one uses the following approximations for the investment model:

$$C_{CAPEX,local[€]} = N_{RRH}(C_{proc[€]} + C_{cab/cell[€]} + C_{IO/cell[€]} + C_{CW/cell[€]} + C_{lic/cell[€]} + C_{FH/cell[€]}) \quad (3.28)$$

where:

- C_{proc} – Cost of processing units (baseband and general units).
- $C_{cab/cell}$ – Cost of baseline components normalised per cell.
- $C_{IO/cell}$ – Cost of fronthaul investment normalised per cell.

$$C_{CAPEX,CRAN[€]} = \sum_{p=1}^{N_{Pools}} N_{RRH,p}(\nu C_{base/cell[€]} + C_{CW/cell[€]} + C_{lic/cell[€]} + C_{FH/cell[€]}) + \lceil N_{RRH,p} V_p \rceil (\rho C_{proc[€]} + C_{IO/cell[€]}) \quad (3.29)$$

where:

- $C_{base/cell}$ – Cost of baseline components in C-RAN normalised per cell (includes switches and controllers).

Regarding OPEX, the following expression accounts for all the components that need to be considered:

$$C_{OPEX[€]} = C_{rent[€]} + C_{power[€]} + C_{mtc[€]} \quad (3.30)$$

where:

- C_{rent} – Rents expenses for equipment housing (per year).
- C_{power} – Power consumption expenses in equipment and in air conditioning (per year).
- C_{mtc} – Maintenance expenses for equipment and transmission network (per year).

The values to study in operational expenses are more unpredictable. For the renting costs, one assumes that cost variations are due to the different total areas in both implementations and different prices per square metre. Note that with this approach, one is considering that mobile operators are negotiating rents based on the square metre, which is not always the case.

$$C_{rent,local[€]} = N_{sites}\, C_{r,local[€/m^2]} A_{cabinet[m^2]} \quad (3.31)$$

$$C_{rent,CRAN[€]} = N_{pools}\, C_{r,CRAN[€/m^2]} A_{pool[m^2]} \quad (3.32)$$

where:

- N_{sites} – Number of sites in the scenario.
- $C_{r,local}$ – Average cost of m^2 for cabinets (per year).
- $C_{r,CRAN}$ – Average cost of m^2 for pools (per year).
- $A_{cabinet}$ – Required area for a cabinet (digital processing component).
- A_{pool} – Required area for a pool.

The power consumption considered in this model is only the one required for digital processing of each

cell (the power for the RRHs is not considered). Significant savings can be achieved by considering that fewer boards are required for processing due to multiplexing gain.

$$C_{power,local[€]} = N_{RHH}\, P_{board[\mathrm{W}]} \frac{365 \times 24_{[h]}}{1000}\, C_{energy[€/\mathrm{kWh}]} \tag{3.33}$$

$$C_{power,CRAN[€]} = \left(\sum_{p=1}^{N_{Pools}} \lceil N_{RRH,p} V_p \rceil \eta P_{board[\mathrm{W}]}\right) \frac{365 \times 24_{[h]}}{1000} C_{energy[€/\mathrm{kWh}]} \tag{3.34}$$

where:

- C_{energy} – Normalised cost of the energy consumed per kWh.
- P_{board} – Power required for one regular board.
- η – Energy efficiency of the boards used in comparison with regular boards.

Maintenance is frequently assigned to external companies and an exact model is not easily defined. One assumes that maintenance corresponds to a fraction of the initial investment.

$$C_{mtc,local[€]} = N_{RRH}(C_{cab/cell[€]} + C_{IO/cell[€]}) l_{cab} + N_{RRH} C_{proc/cell[€]} l_{proc} + N_{RRH} C_{CW/cell[€]} l_{CW} \tag{3.35}$$

$$C_{mtc,CRAN[€]} = N_{RRH}(C_{base/cell[€]} + C_{IO/cell[€]}) l_{cab} + N_{RRH} C_{proc/cell[€]} l_{proc} + N_{RRH} C_{CW/cell[€]} l_{CW} \tag{3.36}$$

where:

- l_x – Share of the investment in component x of the model, spent on its maintenance each year.

3.3 Model Implementation and Algorithms

3.3.1 Generic Algorithm

In this section, the structure of the simulator is described. The inputs required by the program and the outputs exported are addressed. Finally, a detailed explanation of the algorithms used in the simulator is also presented. Figure 3.3 presents the complete generic simulator flowchart.

One can see the inputs that were already commented on Section 3.1.2 and the outputs representing the metrics under study. Between inputs and outputs, one can see the proliferation layer (if required in the input configuration) and the generic flow of RRH-BBU assignment.

3.3.2 RRH Proliferation Module

As mentioned in Section 3.1, it is important to consider how the C-RAN architecture can adapt to the growth of the network justified by the new traffic demands predicted for the upcoming years (Section 1.1). In order to make a more reasonable study of the future C-RAN implementation, and to understand how this architecture can scale with the network, one should consider scaling the overall network traffic per year. The scaling of traffic can be forecasted by increasing the demands in the currently existing base stations, by adding new carriers in the existing base stations or by deploying new ones. The latter

case is applied in this module to simulate the densification of an urban network.

The main goal of this algorithm is to take as input the existing base station locations and traffic patterns and the simulation time in years to introduce new RRHs with different positions and traffic profiles in the scenario under study. The newly deployed RRHs are considered to be low-powered nodes (small cells) to increase the overall capacity of the network as explained in Section 2.3. Another input is the proliferation factor that accounts for the percentage of new RRHs that are expected per year. As the growth rate may not be constant, one uses an adjustable factor for each year of simulation.

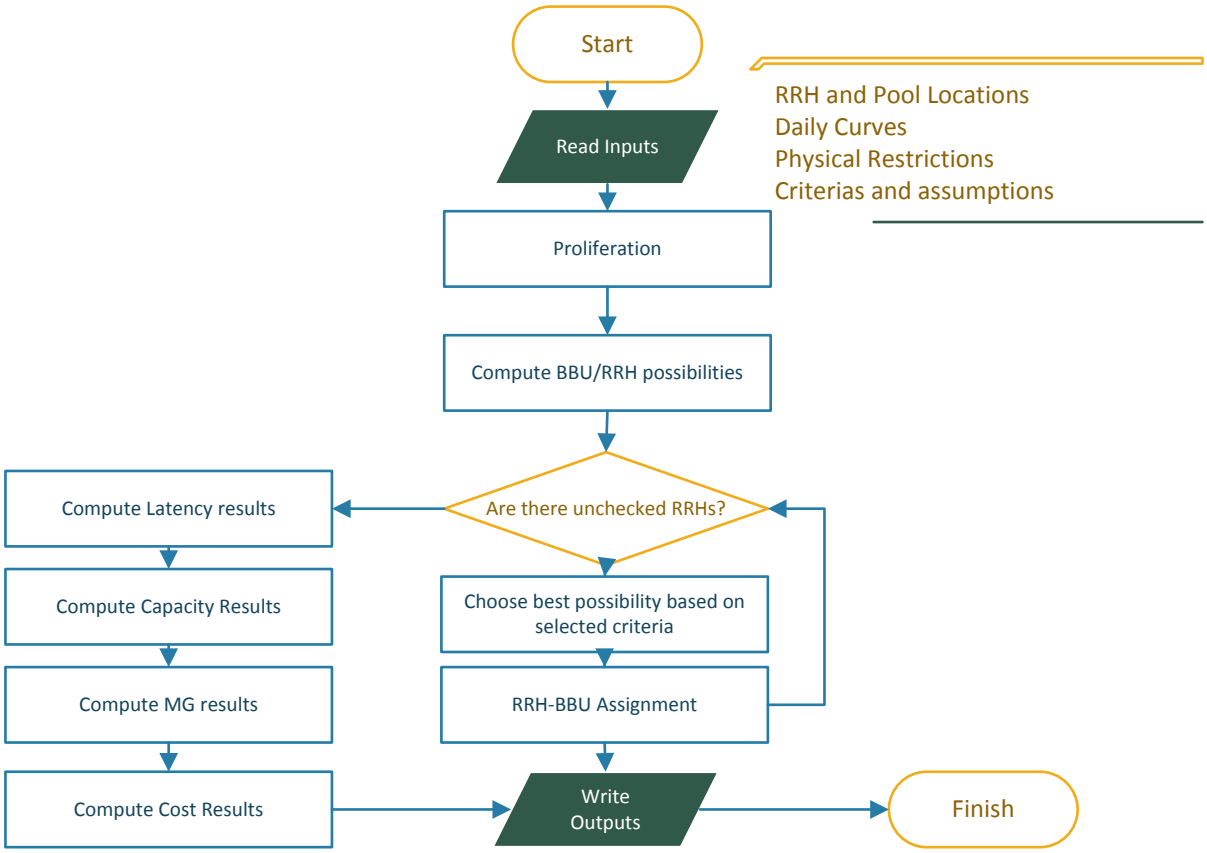

Figure 3.3.Model Implementation Algorithm.

One also considers that this proliferation algorithm based on small cells is relevant for the entire area when all cells in the scenario are dense urban, urban and suburban. Despite this fact, the algorithm takes the different densities of RRHs (base stations per area) to assign more RRHs in the denser zones of networks into consideration. This assignment strategy is chosen because RRH density is usually associated with more populated zones in which traffic is expected to be higher. The algorithm is adapted from [Maro15], working with the Probability Distribution Function of the distances to a reference point in the centre of the scenario. Based on this function, one can extract an element with the probability conditioned by the distance to the reference point. In this way, elements in a denser area have more probability to be selected by the algorithm than the ones in a less dense area.

After selecting one RRH, the algorithm chooses the two nearest neighbours and deploys a new RRH in the centroid of these three points. The newly instantiated RRH is then added to the list of deployed RRHs, the traffic behaviour of that cell being based on one of four different metrics. Figure 3.4 shows

the proliferation algorithm.

To forecast the traffic behaviour of each of the new cells, one uses different approaches representing different proliferation scenarios:

- Traffic average.
- Weighted traffic average.
- Specific traffic profiles based on neighbouring RRHs.
- Specific traffic profiles assigned randomly.

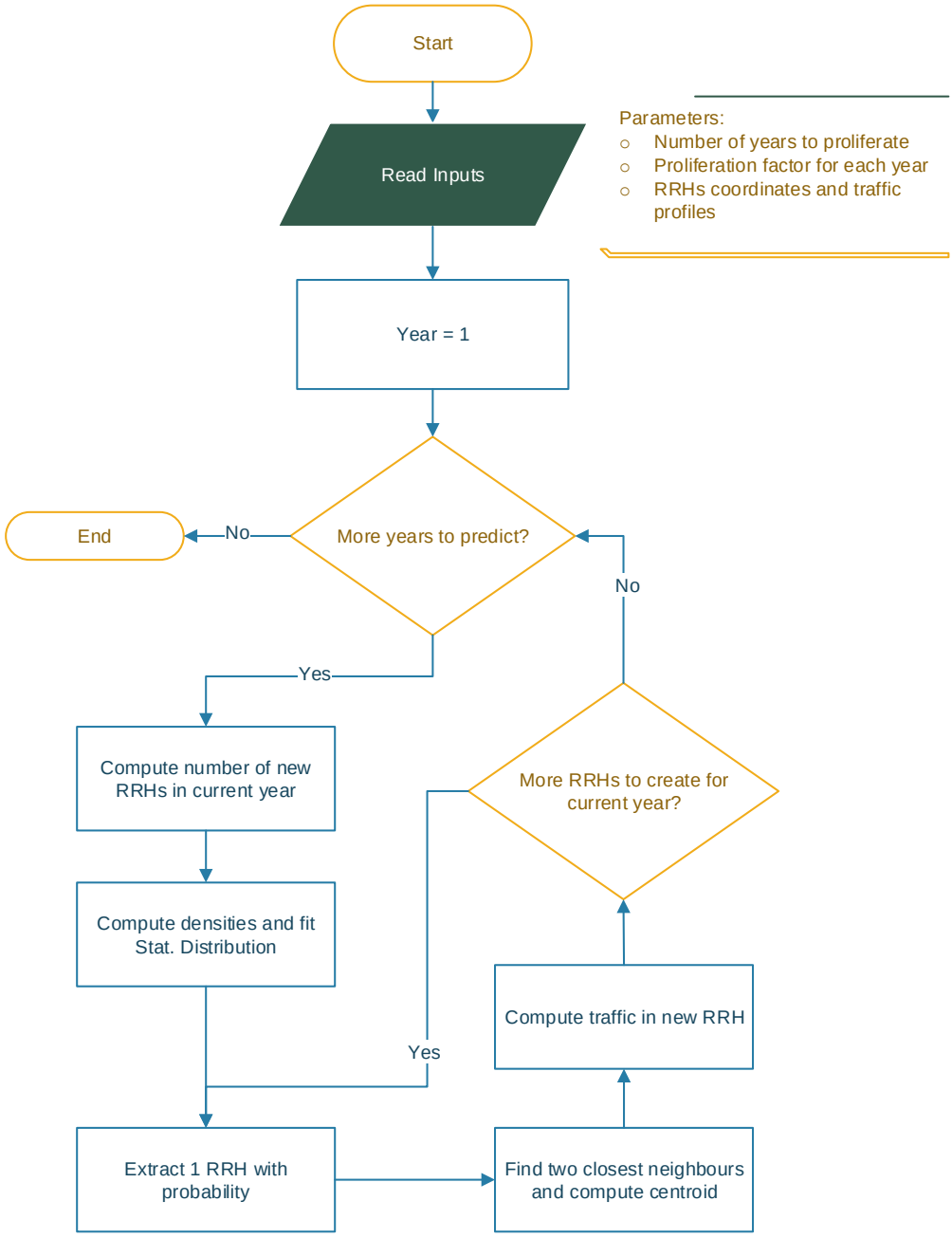

Figure 3.4. Proliferation Module Algorithm.

For **traffic average**, one assumes that the new cells have a traffic profile that reflects the traffic behaviour of the three neighbours used to compute the new location.

With the **weighted traffic average** approach, one attempts to average traffic without considering the neighbours that have a different traffic profile. The main idea behind this algorithm is that small cells are low-powered RRHs that have a short coverage, and that will likely have a very specific traffic profile (deployments in only residential cells or only office cells). The algorithm checks which is the type of cells that is more represented in the neighbouring RRHs, and does the average of the cells that have that classification. If this first criterion is not enough to classify the new cell traffic behaviour, one uses the traffic curve of the cell with higher demand of traffic for the new one.

In the **specific traffic profiles based on neighbouring RRHs**, one uses a different approach without averaging the traffic of the neighbouring cells. The idea is to use specific and fixed traffic profiles for each area. This option attempts to simulate a deployment case in which small cells are placed in the same specific environments, such as shopping centres or residential buildings. To define which of the profiles is assigned to each cell, one uses the same classification metric as in the weighted traffic average approach.

The final approach is **specific traffic profiles assigned randomly**. In this case, the algorithm assigns specific traffic profiles to each cell, independently of neighbouring cells, based on a random decision with different probabilities for each type of cell. This approach is used to study how this deployment ratio can affect the overall gain of clustering multiple RRHs.

3.3.3 Minimise Delay Algorithm

The Minimise Delay algorithm described in Figure 3.5 is designed to assign each RRH to the closest data centre in the scenario, the main goal being to study the minimum values of delay achievable if all RRHs are to be centralised in the entire area. One considers that a centralised cell is a served cell.

The algorithm takes as input the locations of the deployed RRHs and possible locations of BBU pools. To define the list of possible pools to which each RRH can be connected to, a maximum distance is configured based on the maximum delay tolerated by the architecture. If any of the cells is too far of possible pools, it is marked as unserved. One also has the possibility to limit the capacity of each pool, and to mark cells that cannot be centralised due to capacity limits as unserved by capacity. In the latter case, the algorithm will perform better if it is divided into two iterations and if one sorts RRHs by the distance to any pool in order to give preference to the cells that are closer to any data centre. This is valuable mostly if one is also aiming at clustering the neighbouring cells together for CoMP gains.

3.3.4 Load Balancing Algorithm

Another assignment strategy can be the load balancing approach. In this algorithm, one attempts to use all the pools in the scenario to make a balanced distribution of RRHs through all of them. The idea is to have a balanced network that is more resilient to faults due to a higher geographical distribution of base stations. Besides that, a mobile operator can have each pool dimensioned to handle lower traffic than designing a single one to support all the required digital processing. Figure 3.6 shows the algorithm described in this section.

The algorithm takes as input possible pools locations, RRHs coordinates, latency and capacity limits, and the load required for each RRH. Two options are available in the load balancing algorithm. One is for non-virtualisation scenarios in which the main goal is to balance the number of RRHs in each of the pools locations. The other one is for a virtualisation scenario, in which one takes into consideration the load, in MBph or in GOPS, to balance the number of processing cards required in each pool. In the latter case, one should select the time instance in which intends to balance load, but still take into consideration the capacity limits in each hour of the day. The output is the set of connections among RRHs and pools.

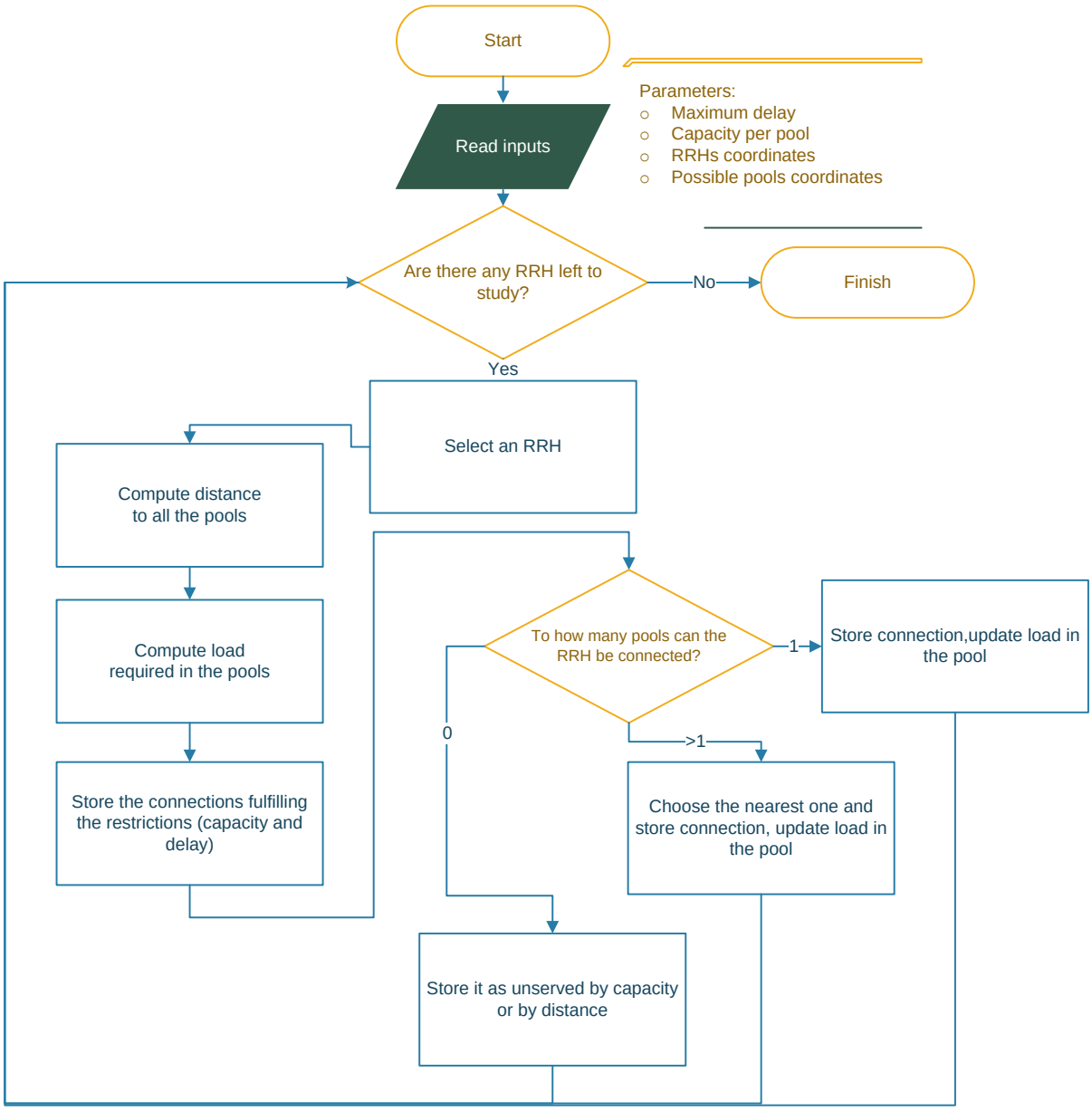

Figure 3.5. Minimise Delay Algorithm.

The algorithm starts by defining a set of connection possibilities for each RRH, being based on the maximum distance (obtained from the maximum delay) allowed by the technology and by the capacity limits of the pools. If one RRH has just one connection possibility, the algorithm assigns the RRH-BBU connection and marks it as served with centralisation. All the RRHs that have multiple possibilities are

studied in the second iteration. In this phase of the algorithm, one assigns each of the RRHs to the pool that has less load allocated to it. The load may be the number of RRHs in the non-virtualisation scenario and the traffic in GBph, or the processing power in GOPS in the virtualisation scenario. In the virtualisation case, one can also sort the RRHs that need to be assigned, based on their capacity in order to prioritise the ones with higher demands that serve more loaded cells. With this approach, a more balanced distribution of load is expected.

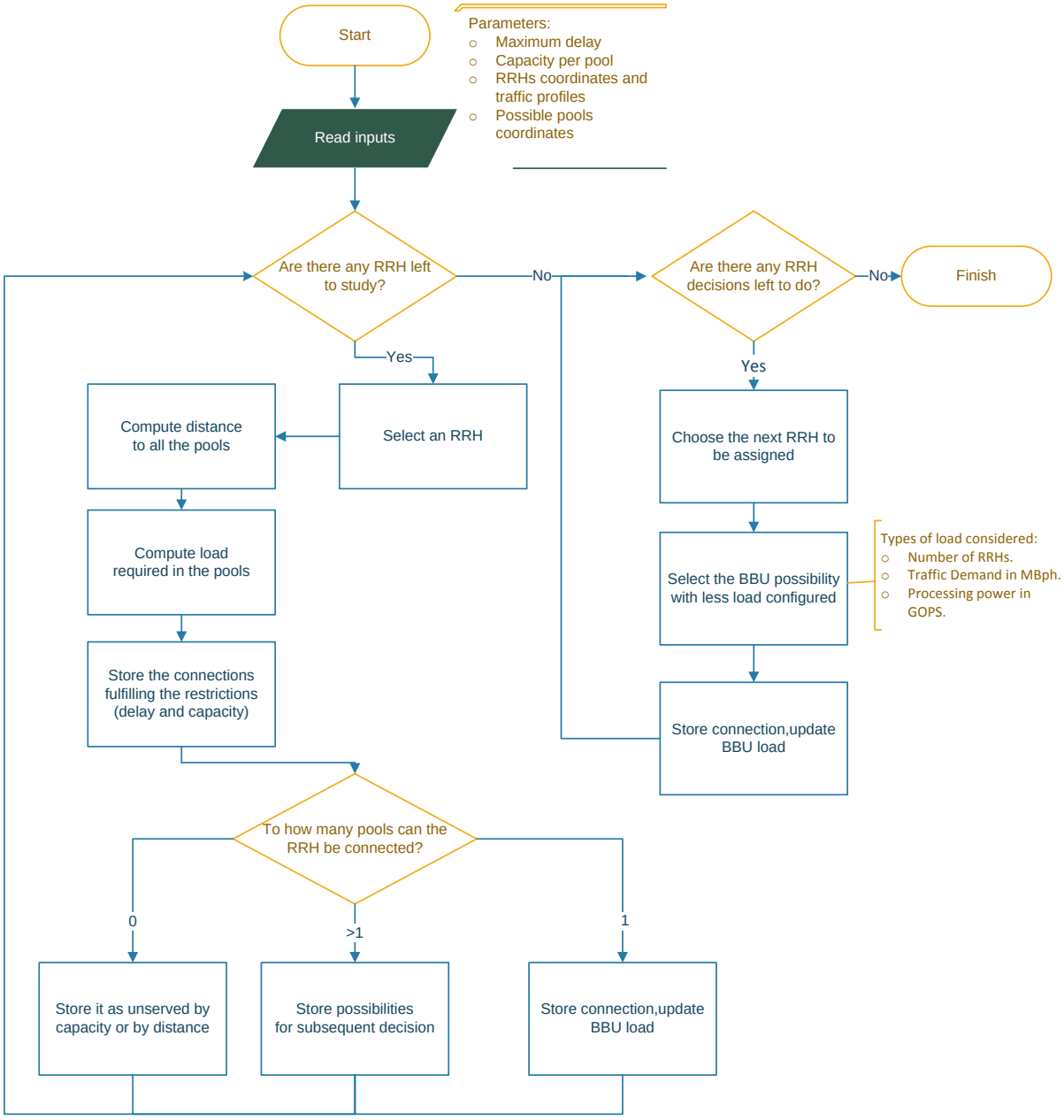

Figure 3.6. Load Balancing Algorithm.

3.3.5 Minimise Number of Pools Algorithm

With this assignment strategy, the idea is to minimise the number of required pools in the scenario. The goal is to maximise centralisation with the assumption that there is no infinite capacity in the pools. The problem is solved based on a heuristic of a known mathematical problem named as vector bin packing

with capacity awareness. The mathematical solution, described in [LKKu99] for parallel processing, is adapted to combine the maximum number of RRHs according to their load requirements in multiple time instances; one uses the sorted Permutation Pack solution proposed by the authors.

The inputs are the RRHs locations, the possible pools locations, the traffic profiles, and capacity and latency restrictions. One uses adjustable time instances (hours of the day) to balance the load through time and maximise resources. The algorithm is presented in Figure 3.7.

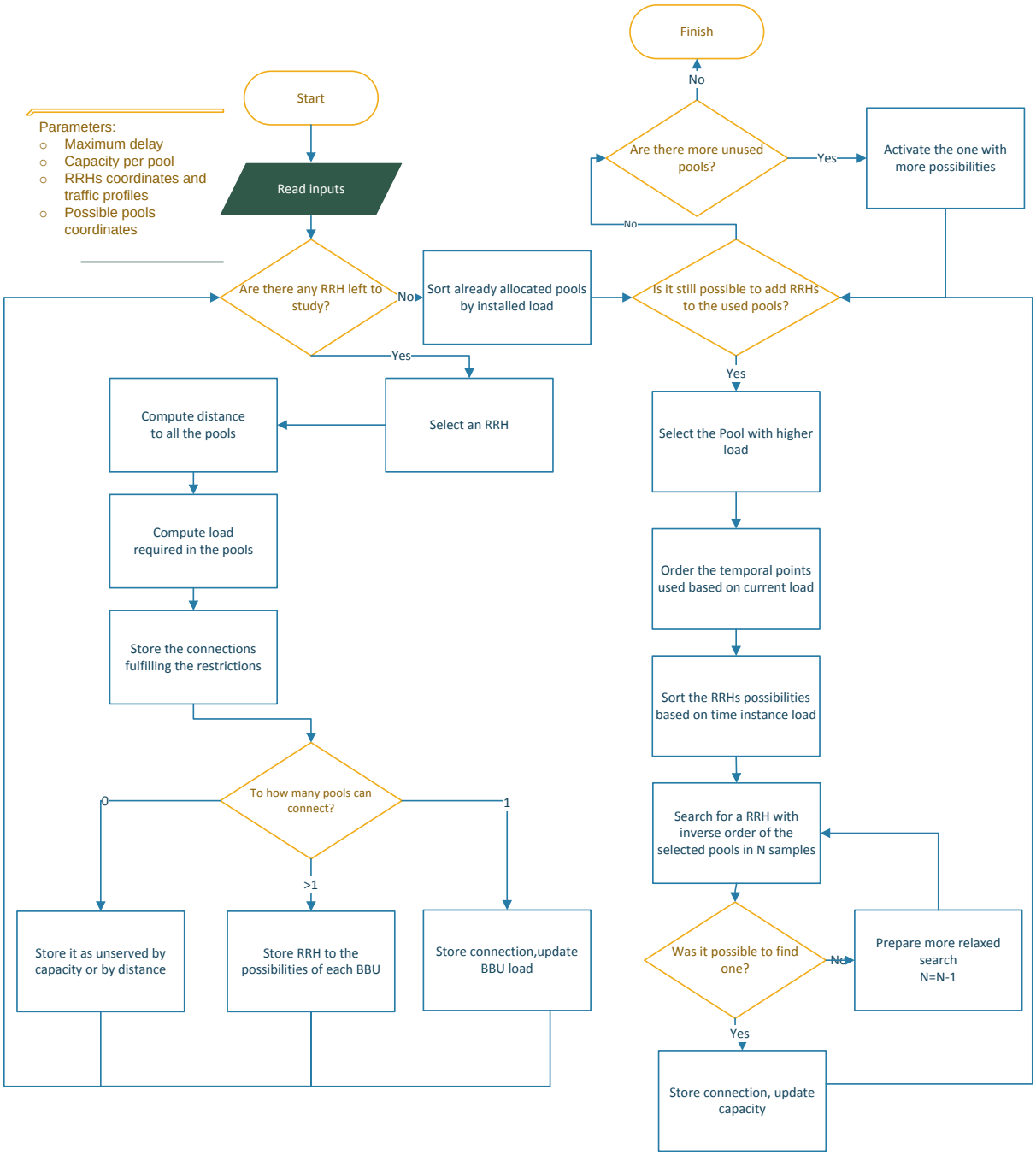

Figure 3.7. Minimise Pools Algorithm.

The algorithm firstly works by assigning the RRHs that only have one possibility. Then, an attempt is made to assign more RRHs to the pools that are already in use. In order to respect the selected time instances, one sorts the load requirements in the pool by hour and searches for an RRH with an inverse

sorted list. If the inverse order is not found in the possible RRHs, one relaxes the matching criteria by searching the same order with one less instance. The latter procedure is repeated until one RRH fulfilling the criteria is found and assigned.

3.3.6 Maximise Multiplexing Gain Algorithm

In this assignment strategy represented in Figure 3.8, one develops an algorithm to force maximum multiplexing gains in the pools.

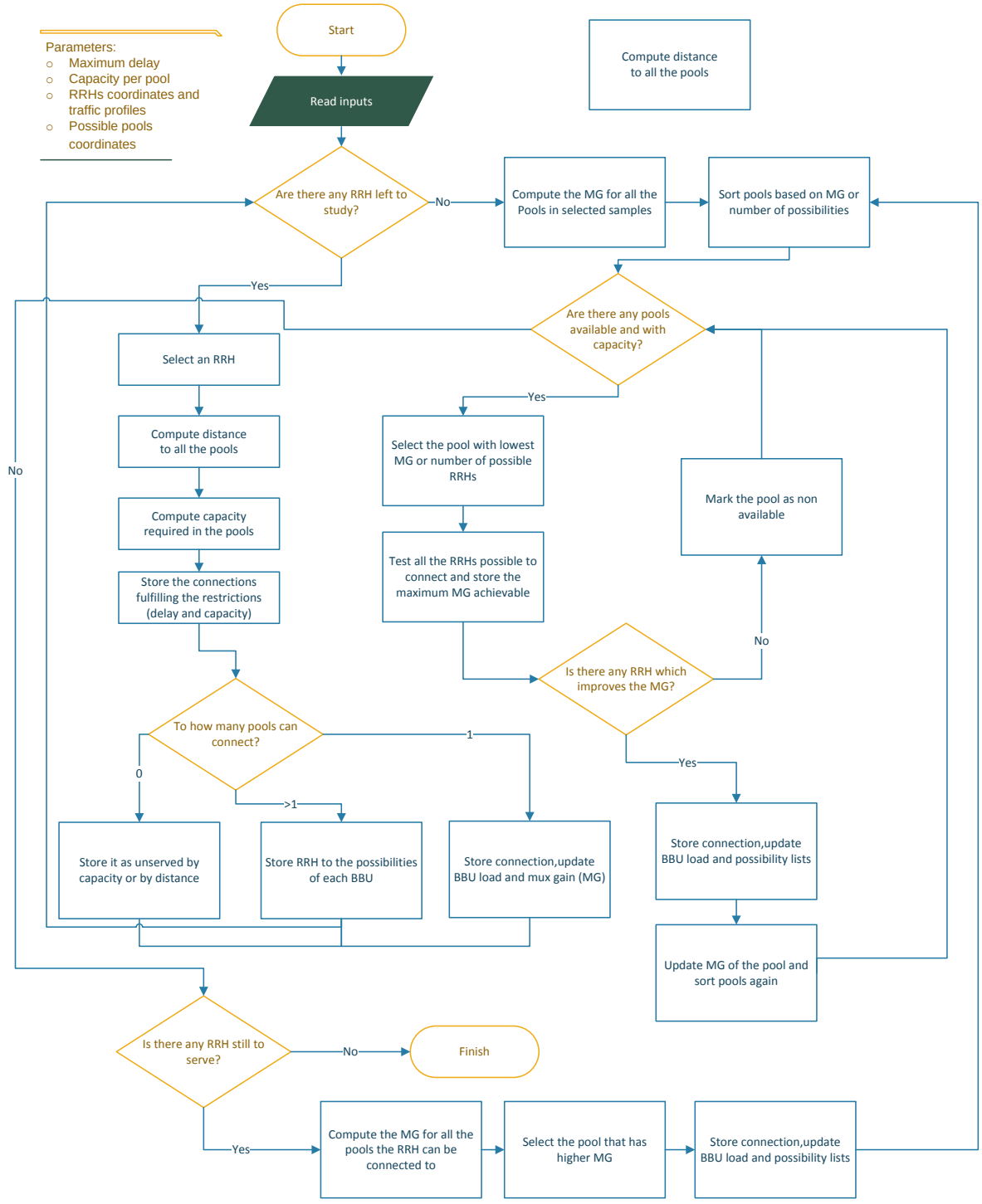

Figure 3.8. Maximise Multiplexing Gain Algorithm.

The idea behind this algorithm is to use all possible pools in the scenario to serve the RRHs, while ensuring that the overall traffic curve of each one is as flat as possible. With this approach, one attempts to correct unbalanced traffic profiles of pools with the RRHs that have a complementary behaviour in time. As a result, the network centralised processing power is lower than in a decentralised network, and lower investment and operational costs are achievable.

The algorithm starts by assigning the RRHs that only have one pool to connect to. A list of RRHs possibilities for each pool is created and updated after each assignment. Then a set of iterations is done in the workflow. In each iteration, one should work with the MG of each set of BBUs. The idea is to assign the RRH that improves the MG to the pool that has a lower gain. If at any instance there are none RRHs improving any of the MGs, one should assign the RRH that has less impact on a pool and starts a new iteration.

The inputs are the RRHs locations, the possible pools locations, the traffic profiles and capacity and latency restrictions. One uses the 24 hours of the day to compute the multiplexing gain.

3.4 Model Assessment

In order to validate model implementation, a set of empirical tests were applied to the simulator to ensure the correct behaviour of the program. The idea of this assessment is to apply a set of tests in which the outcome of the simulator is already expected to see if the outputs of the scripts are accurate from a theoretical viewpoint. In addition, one also tests if the global variables of the scripts are coherent during the execution of the program.

The structural tests that were applied to the program are described in Table 3.2. These tests ensure that the data being read by the simulator are treated properly, and that the multiple assignments in the workflow are executed as expected. Most of the tests were done with Matlab's debugging tool to access the variables values in different phases of the workflow. These tests are particularly important as the core of the simulator is deterministic. The tests were run for the reference scenario used in Section 4.2.

The first logical assessment is to check the percentage of served cells (connected to a pool) in an urban scenario, applying different maximum distances (different delay requirements), Figure 3.9. As expected, increasing the distances implies that more RRHs are served. In the urban scenario under study, a maximum distance of 12 km is enough to ensure that all RRHs have a pool to connect to. This result was already expected, as the maximum RRH-BBU distance in the scenario is 11.4 km. Following the same line of thought, Figure 3.10 represents the same evolution, but with capacity limits of each pool instead of fronthaul limitation. The curve has the same expected trend, as the value of shared cells increases with the increase of traffic handled in an hour by all pools in the scenario. The saturation occurs at a capacity of 17 GBph, this value being very close to the peak of traffic in the network (281 GBph, at 3 p.m.) divided by the number of pools in the scenario (19). It is not the exact value as, the capacity limit is being tested for all the hours of the day, and a geographical distribution is

conditioning load assignment.

Table 3.2. Module Assessment Tests.

Number	Description
1	Validation of the input files read, by verifying if the number of RRHs and BBU pools stored in memory are the same as in the files.
2	Validation of the input variables, by verifying if the simulation criteria and parameters are correctly stored in memory and processed in the workflow.
3	Validation of the RRHs and pools coordinates, by scattering their positions in Matlab and by plotting them over Google Maps and inspecting their placements.
4	Validation of the restrictions in each connection: • Check if the distance between a pool and RRH are below the maximum defined when they are considered as a possible connection. • Check if the load available in a pool is enough to handle the new RRH when they are considered as a possible connection.
5	Validation of the assignments RRH-BBU: • Check if the connection is correctly stored in the pool. • Check if the load in the pool is correctly updated. • Check if the RRH is marked as served and not assigned again.
6	Validation of the assignment algorithms: • Check if the sum of unserved RRHs and served RRHs is the same as the total number of RRHs. • Check if there are not duplicate RRHs in the Pools configurations. • Compare the processing time of each one of them with the expected ones based on computational complexity.
7	Validation of the multidimensional execution when considering multiple time instances in the same simulation: • Check if the load limit is being considered for all the time instances. • Check if the load is being correctly updated in all the time instances.
8	Validation of the output files, by verifying they are printing and plotting the results obtained and already assessed.

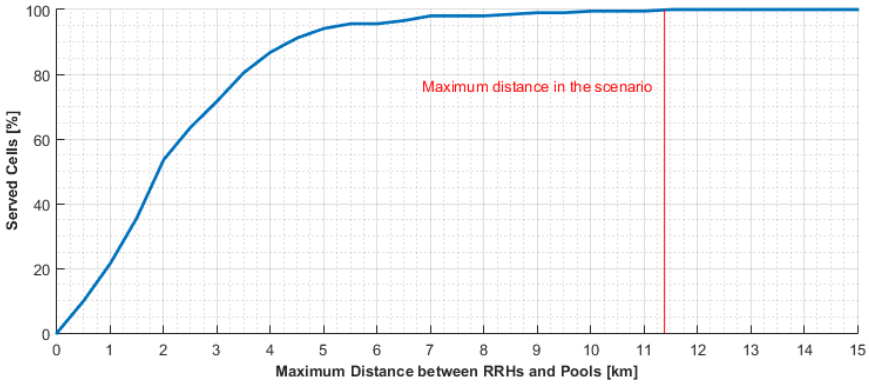

Figure 3.9. Share of centralised cells depending on the maximum distance allowed.

To assess the correct behaviour of the proliferation module, the metric of the density of neighbours (number of neighbouring RRHs per area) is used as well as the total number of sites per year. As this layer is the only one introducing some randomness in the simulator, multiple simulations were done. Figure 3.11 represents the number of sites in each year of simulation. The scenario under study includes 200 initial sites. The number of sites increases each year by a factor corresponding to the proliferation factor, as expected. A radius of 4 km is defined for the assessment in Figure 3.12. The average density of neighbours of all RRHs is presented, as well as the standard deviation. The mean density is obviously

increasing by a factor close to the proliferation factor, 1.3. The standard deviation also increases due to a non-linear proliferation, as explained in Section 3.3.2. These results confirm the non-uniform proliferation that was intended.

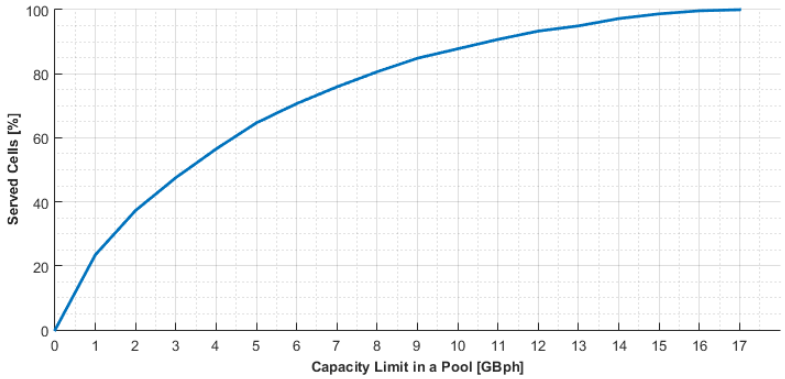

Figure 3.10.Share of centralised cells depending on the capacity imposed in the network.

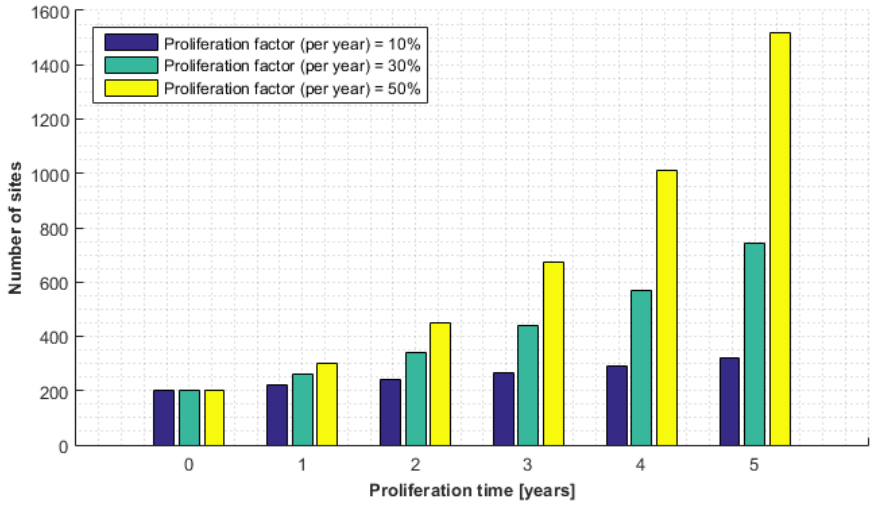

Figure 3.11. Number of sites depending on proliferation time.

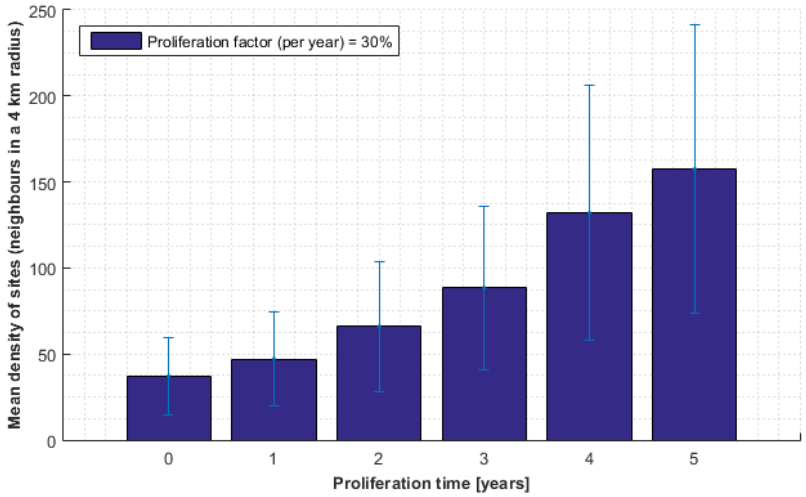

Figure 3.12. Mean density of sites depending on proliferation time.

Another metric used is the program's simulation time. This value is highly dependent on the computer being used, and on the number of RRHs and possible pools used as input. Using an Intel Core i7 2.4 GHz processor with 16 GB of RAM, a scenario with 614 initial RRHs and 19 possible pools, one usually gets processing times in Matlab of less than a minute for the most complex algorithms. Depending on the analysis being done, the outputs being exported can add extra time to process. As expected, the simulation time will increase significantly when proliferation is used, naturally depending on the number of years and the proliferation factors. Using 5 years of proliferation with a constant factor of 1.5 one gets processing time close to 20 minutes. Taking all of these tests into account together with a constant interpretation of results, it is safe to stay that all of the following outcomes are obtained through a fully functional computational tool that implements the desired and described model.

Chapter 4

Results Analysis

This chapter presents the considered scenario, the obtained results, and their analysis. Section 4.1 defines the scenario and the values taken into account for the reference scenario as well as the most relevant assumptions taken. Section 4.2 contains the results of the most important metrics analysed for the reference scenario. All of the sections between Section 4.3 and Section 4.9 represent parameter analysis in which some input value is changed to see the influence on the results, namely latency restriction, uplink traffic, traffic profiles, number of pools, daytime balancing, proliferation and cost analysis.

4.1 Scenarios

This section describes the scenario under study in this thesis.

The geographical region is the city of Porto, particularly the most populated area. The goal is to analyse C-RAN implementation in this city, considered as a dense urban environment. In Figure 4.1, a scenario spanning for 40 km off the city centre in Boavista is presented, where site locations and possible pools are represented as blue triangles and red squares, respectively.

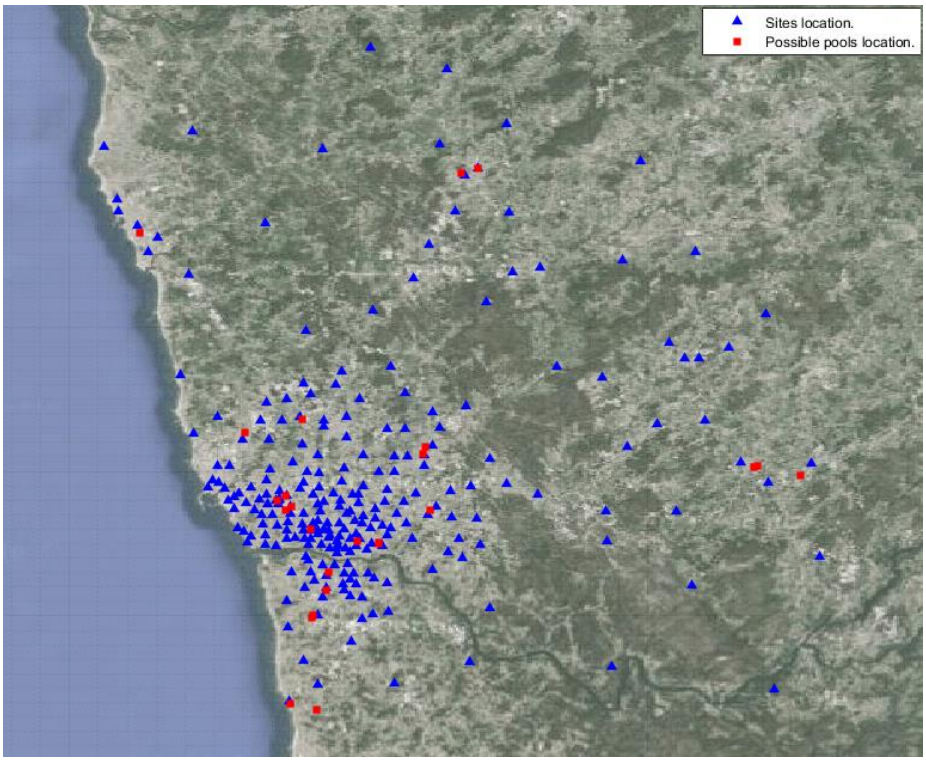

Figure 4.1. Porto scenario with sites and pools location.

Porto is the second biggest city in Portugal. The metropolitan area is constituted by 17 municipalities. Although several of them are presented in Figure 4.1, the ones actually considered for the dense urban zone are mostly Porto, Maia, Matosinhos, Vila Nova de Gaia, Valongo and Gondomar. The density of RRHs in these zones justifies this option. In the described zone there are 1 112 555 inhabitants in an area of 562 km^2 according to the last census report [INEs11].

In the reference scenario, one only considers the dense urban zone of Porto spanning from a radius of 20 km from the city centre. The number of RRHs and BBUs is shown in Table 4.1. It is relevant to note that not all sites have the same number of cells/sectors: although 88% have three, one uses the correct number of cells/sectors in each site for the analysis.

Regarding the RRHs selected configuration, one assumes that all radio equipment is the same to reflect a scenario that is entirely urban. A bandwidth of 20 MHz is used with a MIMO configuration of 2x2.

No capacity fronthaul constraints are assumed. As the main goal is not the fronthaul, one assumes that all RRHs have a fibre network capable of connecting it to any pool. Those fibre links have the available bandwidth for the RRHs configurations and have switching schemes that allow reconfiguring the RRH-BBU connections. [Silv16] contains a more in-depth study of fronthaul, not only with other options other than fibre, but also the impact that fibre deployment would have.

Table 4.1. Number of elements in the scenario.

Components	Number
Sites	200
Cells (RRHs)	614
Possible Pools	19

A value of $2x10^8$ m/s is used for the propagation speed in fibre. The maximum distance considered is 15 km, corresponding to a latency requirement of 150 µs (round trip time), being used to enforce a stringent value and comply with LTE Advance technologies that are one of the biggest benefits of C-RAN deployment in urban environments. Note that, in reality, RRH-BBU links are not straight lines (having a higher delay) and that the deployment of switching schemes introduces more latency due to more equipment and more fibre length. With that being said, the overall value assumed is close to the 200 µs used in literature.

One assumes that technological implementations with resources sharing are available. Although the virtualisation and pooling in C-RAN are still an area under study, one considers this deployment as a reference in order to analyse the full benefits of the centralisation concept. One uses the traffic measures in GBph and GOPS in order to perform load balancing and to analyse the multiplexing gain. In the load balancing case, the reference hour is 10 p.m., due to the peak traffic in the network. DL LTE traffic is used as this link produces more traffic than UL, as expected. In the case of the maximisation of MG, one uses the 24h of the day as time samples. There is no limit to the cluster size, meaning that all cells connected to the same pool can communicate with each other and introduce CoMP gains in the network. As the real values for the processing power capacity are not known, one has used simple rules to adapt the traffic per hour curve to a daily processing power variation. The spectral efficiency parameter can have the closest integer value between 1 and 6 resulting from the normalisation of peak of traffic of the cell against the most loaded cell of the scenario (in GBph). The load varies during the day with the normalisation against the peak traffic of that cell (corresponding to 80% of load). In this way, the scenario is simulated to reflect cells in different radio conditions or type of services demands, and variations of the load during the day.

In reference simulations, the capacity used for the pools is infinite, as more processing cards can be added to the pools if needed.

Finally, the costs reference cost assumptions are summarised in Table 4.2. In this case, one also considers a local solution in order to provide a cost comparison between the two approaches. For the processing cards cost, one assumes that the value is the same for all the analyses (generic boards). The licences costs are being presented together with the baseline unit values. The licences are paid in a “pay-as-you-grow” manner. For model simplicity, one assumes an average of 100 users per cell and

the price of those licences. The energy cost was taken as the average value paid in Portugal during busy hours. The power consumed by a board is based on the reference value of the currently available BBUs and a tolerance is added to account for other equipment consuming power. The price of renting is based on the values usually paid per m^2 in the city of Porto. One uses a smaller value for central pools location, because their positioning is more flexible and can be made according to price considerations. The maintenance factors are derived due to the depreciation of computers (or boards), for the eventuality of civil repairs and regular hardware depreciation (interfaces, switches and power units). The virtualisation factor used as a reference is the inverse of the multiplexing gain measured for traffic curves in GBph. This assumption is that users can be assigned dynamically trough the available boards in a fully load dependent implementation. It is also important to note that this assumption considers that the same gain is achievable in the UL, as the value obtained only concerns the DL scenario. Using the same approach, one considers the available additional complexity factors as 1. It is relevant to state that reference values for those values are not well-defined, being hugely dependent on technical implementation. For that reason, those values are hard to quantify.

Table 4.2. Reference values for cost layer.

Cost Components	Local	Reference CRAN
$C_{proc}[€]$	600	
$C_{cab/cell} + C_{lic/cell}[€]$	300	
$C_{base/cell} + C_{lic/cell}[€]$		450
$C_{CW/site}[€]$	8000	1000
$C_{IO/cell}[€]$	170	
$C_{FH/cell}[€]$		
η		1
ν		1
ρ		1
V_p		$\frac{1}{G_{mux,p}}$
$C_{energy}[€/\mathbf{kWh}]$	0.16	
$l_{base}[\%]$	10	
$l_{proc}[\%]$	25	
$l_{CW}[\%]$	1	
$P_{board}[W]$	60	
$C_{m^2,local}[€]$	10x12	
$C_{m^2,CRAN}[€]$		8x12
$A_{cabinet}[\mathbf{m^2}]$	4	
$A_{pool}[\mathbf{m^2}]$		25

As described in Section 3.2.4, multiplexing gains are only possible due to the existence of different types of cells depending on their traffic demand curves. To properly characterise the scenario based on the type of cells, one uses the metric in Annex C applied to traffic profiles to classify the RRHs as office cells, residential cells or mixed cells. Table 4.3 shows the ratios of each type represented in the picture.

The data used is the real average traffic demand in GBph in each one of the 200 sites. To use a cell assignment strategy, one assumes that each of the cells of the site shares the same traffic as the other

ones.

Table 4.3. Scenario classification based on traffic profiles.

Type of Cell	Office	Residential	Mixed
Ratio of cells in the scenario [%]	37.1	43.2	19.7

Figure 4.2 presents the reference scenario with site classification. Annex D contains the average values of the traffic curves reflecting the profiles of each one of them. One can see that the classification system used is generating traffic curves as the ones presented in [CMRI11]. However, one can see that this scenario contains about 25% of cells that do not have a traffic curve with significant daily traffic variations, a case that was not considered by the authors. This type of cell reflects a realistic scenario of cells with an area that includes both residential zones and offices. In fact, one can see the mixed cells in the most external zones of the scenario due to higher coverage zones spanning through the heterogeneous zone of traffic.

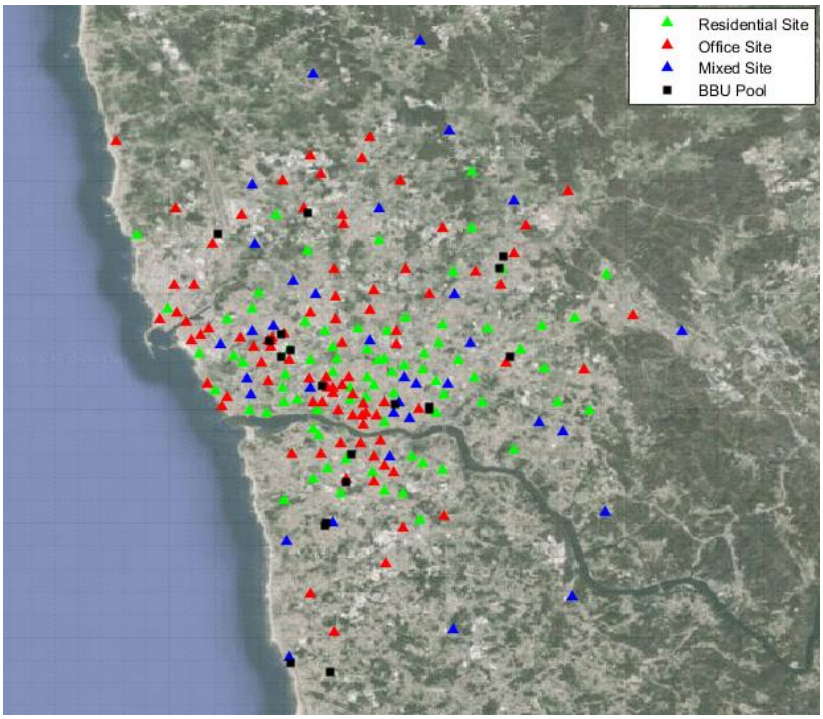

Figure 4.2. Reference scenario with site classification.

4.2 Reference Scenario Analysis

This section is used to present the results obtained in the reference scenario. The section is also used to perform algorithm comparison and to define which one should be used in the later analysis.

The first parameter to be studied is latency. Although the reference scenario has a dimension comparable to the maximum link distance (20 and 15 km, respectively), and the 19 pools are spread along the area, one should measure delay values and use them as a reference for future technologies.

Figure 4.3 presents the distance results obtained under the conditions described in Section 4.1. All RRHs can be centralised under the delay restriction. The results were obtained using the Minimise Delay, the Load Balancing and the Maximise Multiplexing Gain algorithms (Section 3.3.3). As one can see, all RRHs have a link with distance below 15 km; in terms of delay, these values are easily convertible with (3.3). Table 4.4 summarises the latency values for the reference scenario with minimising delay.

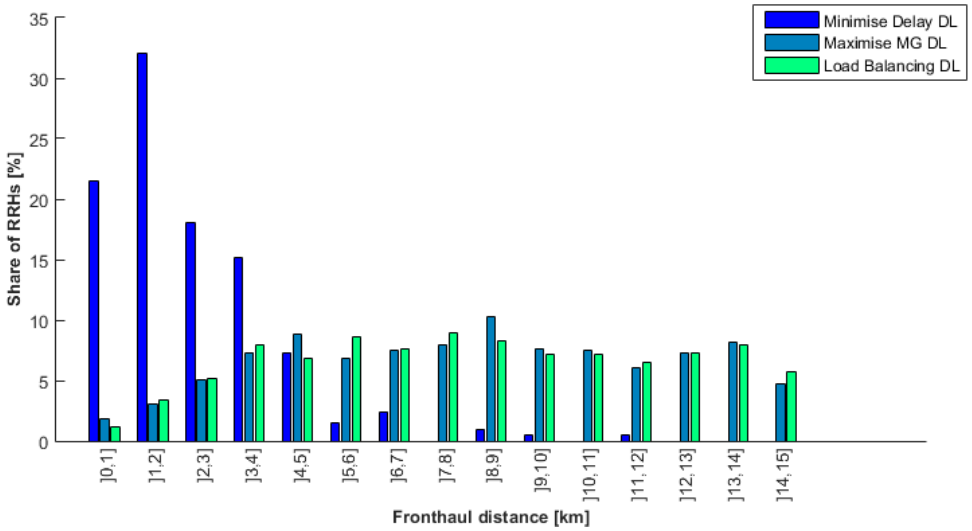

Figure 4.3. Share of RRHs with FH distance (algorithm comparison).

As it is noticeable, the Minimise Delay algorithm has shorter fronthaul distances than the others, as expected, due to the fact that load balancing and multiplexing gain strategies making use of the maximum allowed fronthaul to achieve different metrics other than delay. As expected, the heterogeneity of traffic curves along the scenario introduces larger fronthaul connections respecting the delay restriction. In both scenarios, a considerable number of connections (around 5%) have a length of more than 14 km.

Table 4.4. Distance and latency values for reference scenario (algorithm comparison).

Algorithm	**Latency (two-way) [µs]**	
	µ	**σ**
Minimise Delay	23.11	17.56
Load Balancing at 10 p.m.	81.32	38.28
Maximise MG	81.41	37.58

Concerning the average capacity based on the average values of traffic per hour, Figure 4.4 shows the results of the three algorithms. The Load Balancing algorithm is run at 10 p.m. Naturally, this one is aiming at a more balanced distribution of capacity as the time sample under study is one of the most loaded during the day. With the Maximise MG algorithm, one actually achieves a balance that is close to the one obtained with load balancing. In fact, with maximising MG algorithm, one is actually using a balanced approach to obtain higher gains by assigning cells over all possibilities. As the values of traffic profiles under study are already a result of an average, one should notice that this actual capacity of the network should be higher to handle traffic fluctuations during the days. Table 4.5 resumes this analysis

with the average and standard deviation obtained. One can see the standard deviation value supporting the results of the balancing algorithm. Note that significant variations of load, such as the ones obtained in minimising delay algorithm, mean that an unfair distribution of cells per pools is occurring. In fact, in this algorithm, one gets pools with 85 more RRHs than other. However, in the balanced approach, one gets a maximum difference of only 6 pools.

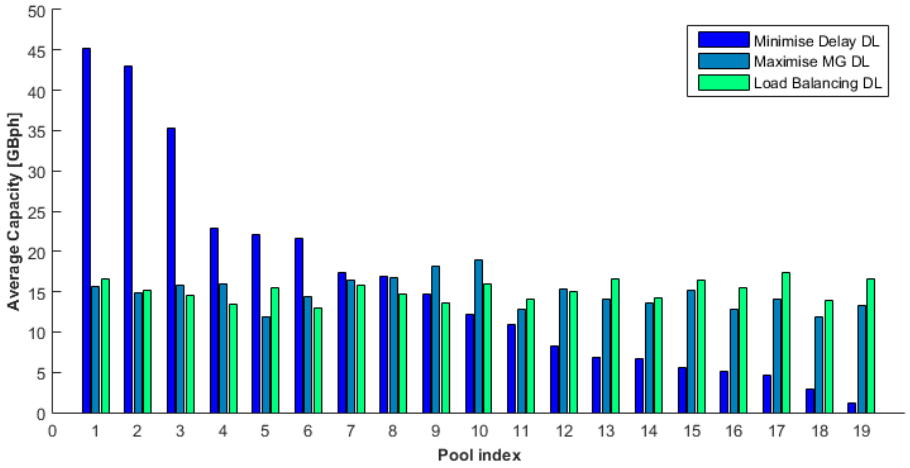

Figure 4.4. Average capacity by each pool (algorithm comparison).

Table 4.5. Average capacity results (algorithm comparison).

Algorithm	Average Capacity [GBph]			
	μ	σ	Maximum	Minimum
Minimise Delay	15.96	13.10	45.19	1.11
Load Balancing	15.16	0.57	17.33	13.02
Maximise MG	14.83	1.93	18.89	11.87

In order to analyse the multiplexing gain, one uses the same approach as the average capacity. Figure 4.5 presents the results obtained for each pool.

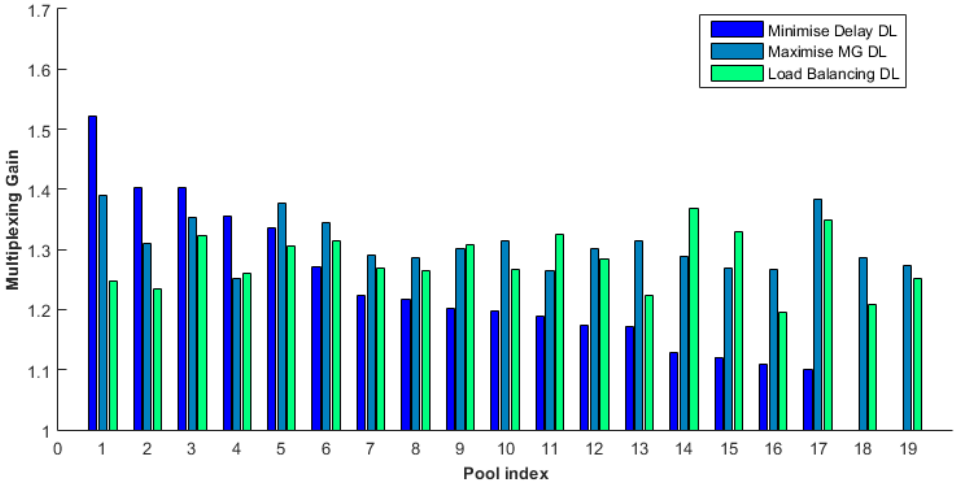

Figure 4.5. Multiplexing gain in each pool (algorithm comparison).

The peak gain is actually achieved in the Minimise Delay algorithm. However, in this last approach, one

actually gets some pools with low MG values, and the overall gain of the network is quite under the other two algorithms. In the Load Balancing strategy and in the Maximise MG, one can see a much more balanced MG throughout the scenario. Nevertheless, the most representative value is the overall gain, which represents the amount of capacity that can be reduced with centralisation. In Table 4.6, one can see the overall gain of the network. One can also the mean and standard deviation of the MG of each pool. The MG value of the Maximise algorithm is 3% higher than the one obtained with Load Balancing. Besides proving that the algorithm achieves higher gains, the latest fact also supports that both produce comparable results in what concerns balancing and gain.

Table 4.6. Multiplexing gain results based on traffic (algorithm comparison).

Algorithm	Multiplexing Gain				
	Network	μ	σ	Maximum	Minimum
Minimise Delay	1.22	1.22	0.14	1.52	1.00
Load Balancing	1.28	1.28	0.05	1.37	1.20
Maximise MG	1.31	1.31	0.04	1.39	1.25

As already discussed, the algorithms are being based on traffic to use realistic data. One also presents the results obtained with the GOPS scenario applied. Although data is not the measured one, the values are also very relevant for current applications of LTE technologies, as the MG application affects mostly the computational processing power. Table 4.7 presents the values obtained in multiple algorithms, but now using the other two gains identified in Section 3.2.4.

Table 4.7. Multiplexing gain types comparison.

Algorithm	Gain		
	Multiplexing	Processing	Processing with under provisioning
Minimise Delay	1.215	1.024	1.270
Load Balancing	1.279	1.032	1.276
Maximise MG	1.308	1.033	1.278

The first thing to notice is that the MG measured in processing capacity is quite low compared to the traffic one. This occurs mostly because the great majority of computational functions are not load-dependent, as it is shown in Section 3.2.3. With that being said, it is expected to have a small variation of resources throughout the day. Besides that, the large variation in processing power occurs due to different radio conditions and not to different traffic profiles. The second thing to notice is that the variation among algorithms is much less significant, which may justify the use of an algorithm that maximises metrics other than MG. When considering an approach with under-provisioning, one can see much higher gains. Although this value is close to the multiplexing gain with traffic, one should note that only a small portion is due to centralisation of resources. The under-provisioning is not specific of C-RAN, but it is certainly facilitated due to the efficient scalability of data centres; the mobile operator can then prepare their processing power with a certain margin and upgrade them when required.

Considering the load analysis based on GOPS, one also gets the results of computational processing power. Figure 4.6 shows the results obtained under the Load Balancing algorithm. The difference between the peak processing power in a day and the lowest one is not as significant as traffic variation,

which supports the low values obtained for MG. The effectiveness of the balancing approach is also noticeable as all the pools process around 1.9 Tera Operations Per Second (TOPS). One can still add that the Maximise MG and the Load Balancing algorithms are not so close in terms of balancing as in traffic per hour, as the values presented in Table 4.8 confirm.

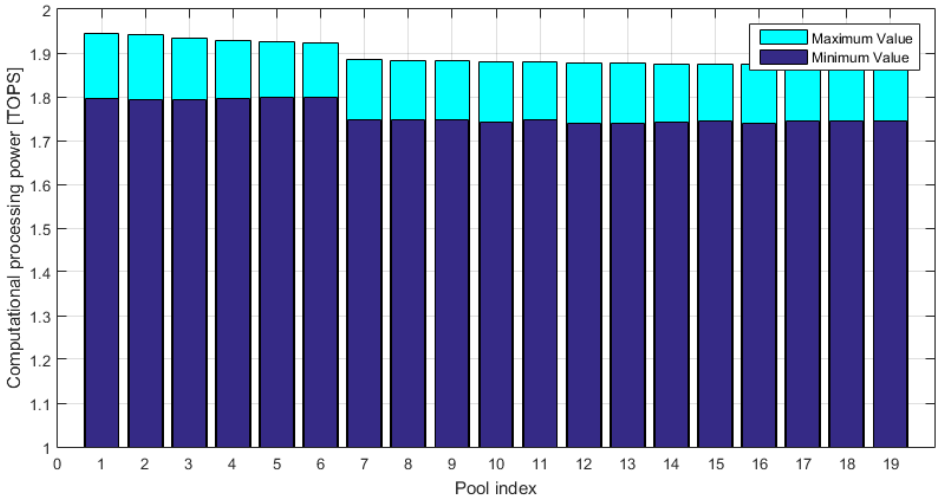

Figure 4.6.Computational processing power per pool (Load Balancing algorithm).

Table 4.8. Average processing capacity results (algorithm comparison).

Algorithm	Average Capacity [TOPS]			
	μ	σ	Maximum	Minimum
Load Balancing	1.89	0.03	1.94	1.87
Maximise MG	1.89	0.22	2.23	1.46

Finally, to analyse the C-RAN deployment one plots the cost achieved with the adoption of the technology. One uses the Maximise MG algorithm, as it is the one with a higher gain and higher cost savings. As expected, there is a considerable cost reduction in CAPEX and OPEX, Figure 4.7.

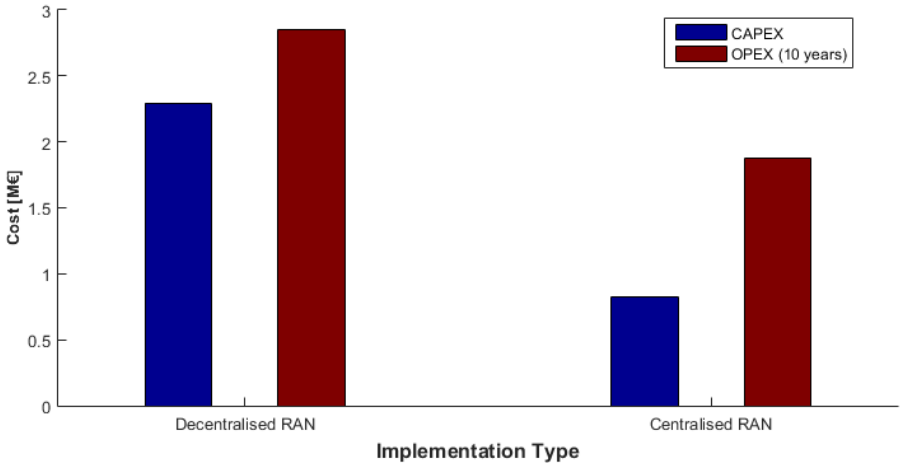

Figure 4.7. CAPEX and OPEX comparison (in 10 years).

With centralisation, one can see that there is about 63% of savings in CAPEX and 31% in OPEX. These

values are proving that there are very considerable advantages of the C-RAN technology compared to another green-field investment in RRH-BBU local technology. One should only note that a full green-field investment is not always the case for a mobile operator. Figure 4.8 and Figure 4.9 present the distribution of cost through the multiple components of the model in CAPEX and OPEX, respectively.

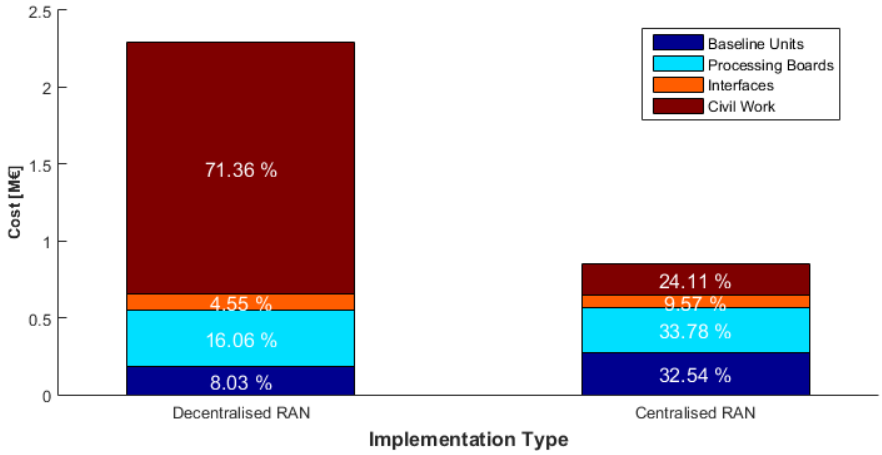

Figure 4.8. CAPEX cost distribution.

Figure 4.8 proves that the main cost reduction is not from the multiplexing gain, but from the civil work involved (around 88% of total reduction). Naturally, the savings in interfaces, boards and civil work can justify a larger investment in the auxiliary technology and licences that increases 50%. However, the savings in hardware only are not enough pronounced to solely justify a C-RAN investment.

Regarding OPEX, Figure 4.9 shows that renting is the main source of savings in operational expenditures with a reduction of 60% of the decentralised cost, although some important reduction is seen in both energy and maintenance (about 20% of savings in each one), as expected, since there are less and cheaper areas to rent when considering the locations for data centres. Energy consumption has savings due to the multiplexing gain, and one can see a large influence on the value obtained in this component. As the hardware does not have such a variation in C-RAN, the maintenance expenditures savings are not so pronounced. Still, one can see a reduction that reflects the easier maintenance of a few pool instead of hundreds of complete sites.

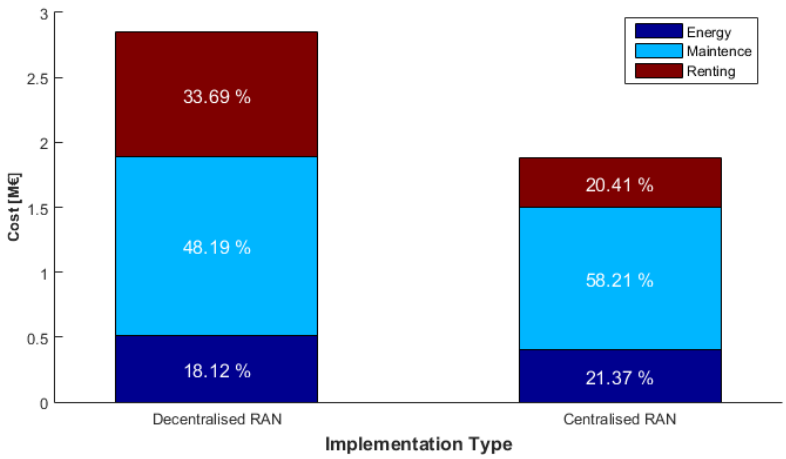

Figure 4.9. OPEX cost distribution (10 years).

4.3 Latency Analysis

In order to fully understand the effect of the maximum fronthaul distance that is considered in the algorithms, one changed the input values of the restriction and analyses multiple outputs: the values are all under the 15 km reference, as this restriction is always expected to be more stringent with the adoption of new technologies, and with the use of extra equipment introducing additional latency (switches, GPPs, etc.). As seen in Figure 4.10, the share of traffic load being handled in a pool is not always 100%, as some of the RRHs do not have a possible pool within the maximum fronthaul distance. With that being said, one actually is studying the effect of partial centralisation in the scenario under different delay constraints.

Naturally, with more relaxed requirements one gets more flexibility in the algorithms being considered. It is not natural to analyse load distribution in the more stringent values of latency, as most of the share of the load is not centralised. For that reason, one plots only the evolution of the multiplexing gain with the Maximise MG algorithm in Figure 4.11.

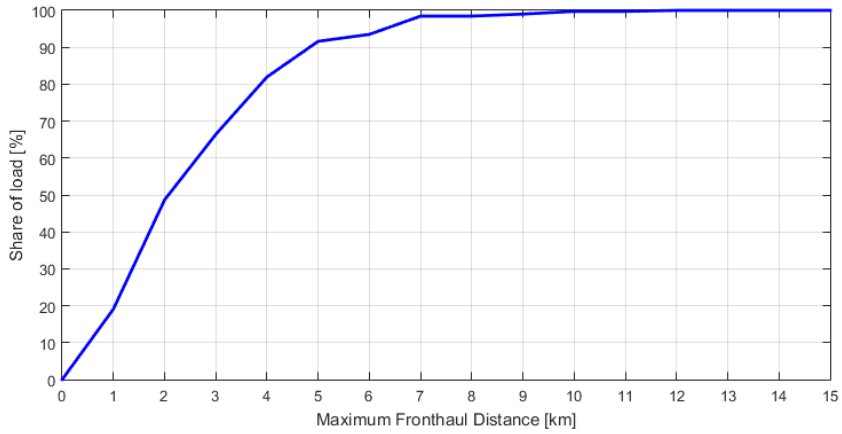

Figure 4.10. Share of traffic load being treated in a pool with maximum fronthaul distance.

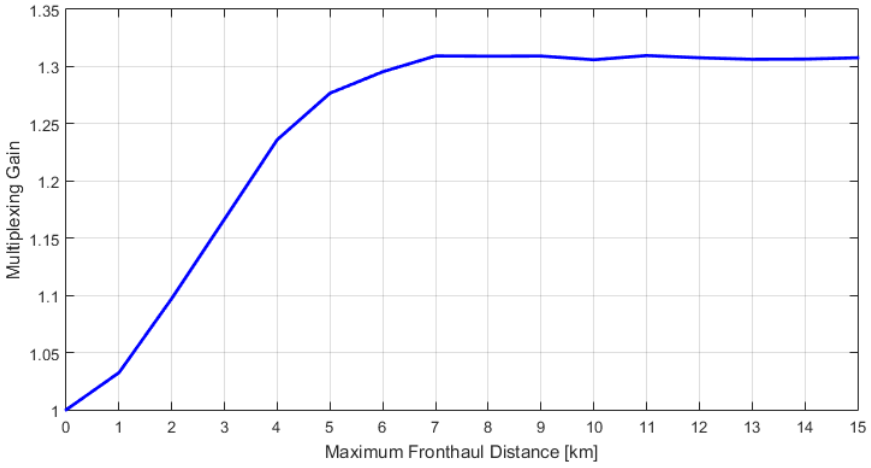

Figure 4.11. Multiplexing gain with maximum fronthaul distance (Maximise MG algorithm).

The multiplexing gain starts increasing as more RRHs are served in a centralised pool. As expected, decentralised BBUs do not contribute to the overall gain. Another interesting remark is that the overall

gain stops increasing for fronthaul limits larger than 7 km, above which one can even note that not all of the pools are centralised (around 2%). The reason for this result is that the cells to be served are the peripheral ones, and are mostly characterised by mixed traffic. As previously commented, mixed traffic cells are the ones that do not improve multiplexing gain as their profile is already flat throughout the day. When all cells are centralised, the MG continues to be around the same value. Although one could have expected an increased gain due to more flexibility in the algorithm, one should note that the algorithm was designed with a balancing approach. With that being said, the strategy implemented is forcing the MG to be balanced in all pools and for that reason, the overall gain of the network is not improving beyond 1.31.

To conclude the analysis on the centralisation achieved with different delay requirements, one analyses the cost, CAPEX and OPEX, evolution for the different distance values, Figure 4.12.

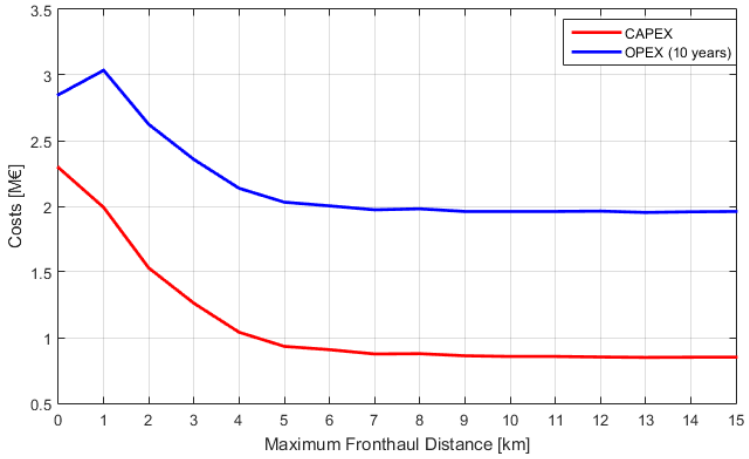

Figure 4.12. TCO with maximum fronthaul distance (Maximise MG algorithm).

The case with fronthaul distance of 0 km is the fully decentralised approach, and over 12 km is the fully centralised one. The declining region of both curves is explained by the natural centralisation of resources with the major contributions coming from the civil work component in the CAPEX, and by the increase of the MG and all the components of operational expenses in the same order of magnitude. Note that OPEX registers an increase with partial centralisation that is explained by the high expenses of renting both sites and pools that are not being properly utilised. The constant values obtained with distances over 7 km are justified, as one as already saw that the MG is constant and that 98% of the load is already centralised. Small variations are only due to the algorithm and different starting points as the pools start getting more RRHs to choose from.

4.4 Uplink Analysis

This analysis is intended to compare the results obtained with UL and DL traffics. The parameter to be changed is the traffic profile of each of the RRHs. As this is the only variation in this section, one should

analyse the traffic related parameters in the pool. Table 4.9 shows the values obtained for traffic in UL. Comparing with the previous reference results in DL, Table 4.7, one can see that the values from all algorithms under consideration are significantly higher, proving that UL profiles are more heterogeneous and that a higher gain is achieved. However, one should also note that the traffic being considered in this approach is lower than the DL one. To proceed with the same comparison, one also studies the effects of gain in computational power. Processing gains with and without provisioning are also shown. The first conclusion is that processing gains are higher; besides the more specific traffic curves explained before, UL processing power is more load dependent and variations have more impact. Finally, the results for under-provisioning prove that much of the RRHs are processing significantly under the maximum processing capacity.

Table 4.9. Multiplexing gain results based on UL traffic (algorithm comparison).

Algorithm	Gain		
	Multiplexing	Processing	Processing with under-provisioning
Minimise Delay	1.337	1.174	1.680
Load Balancing	1.474	1.220	1.748
Maximise MG	1.542	1.233	1.764

To compare the traffic and processing power in each direction, one plots the mean values of the load per pool. Figure 4.13 shows the results for traffic per hour with the algorithm of Load Balancing working for each one of the directions. As one can see, UL traffic is around seven times under DL one.

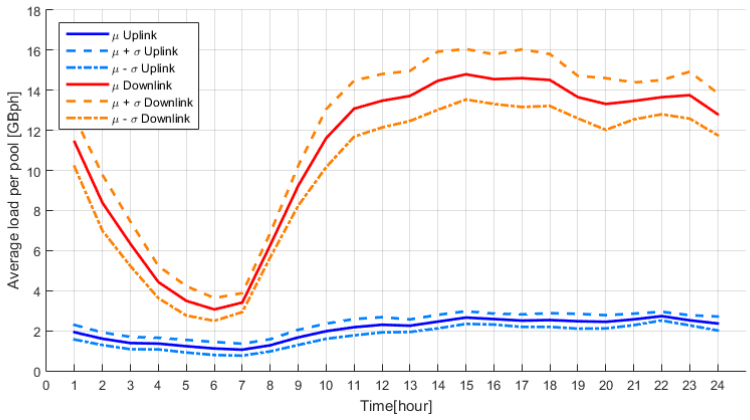

Figure 4.13. Average traffic in each pool within the day.

This conclusion supports the initial approach of considering the multiplexing gain of DL as the one used in the cost model, as the total traffic is quite above the opposite link. However, UL traffic is increasing each year, due to multiple services, such as streaming or social networks, and the results for UL should be taken into account. If one analyses the processing capacity plotted in Figure 4.14, one can see that UL processing power is higher than DL one. This conclusion comes from the fact that UL processing functions are more complex, due to radio characteristics, and require more operation per second in each RRH. The relevance of this result implies that UL should be considered in processing power gain analysis. Nevertheless, one should have in mind that the algorithms under analysis are being optimised

just for one link, so the results displayed are not necessarily for the same network configuration. Is it still clear from Figure 4.14 that UL processing power is much more load dependent than DL's, as DL profile is almost flat during the day, opposite to UL's. As UL introduces higher gains, the previous conclusions support optimistic results in processing gains (computational resources) despite those obtained for DL.

Both figures also sustain previous results, since the standard deviation of average traffic at 10 p.m. is the lowest during the day, as expected due to load balancing. The effect that load balancing at 10 p.m. has on the flat average traffic profile is apparent, justifying multiplexing gains close to the ones obtained with maximising MG algorithm.

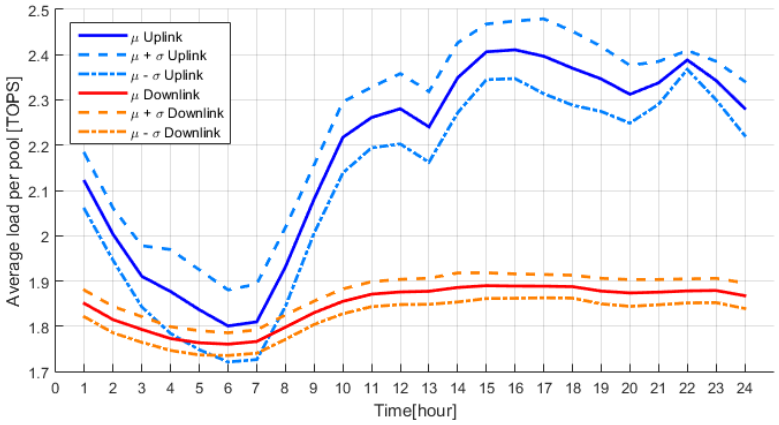

Figure 4.14. Average processing power in each pool within the day.

4.5 Traffic Profiles Analysis

Another parameter under study is the influence of the traffic profiles. Complementing the results obtained in Section 4.4, one uses now reference values from other RATs such as 2G and 3G. The idea behind this problem is to consider all the traffic being handled by the network in a futuristic approach of combining all technologies in centralised pools. Figure 4.15 represents the average load per pool obtained with all traffic profiles. The used algorithm was the Maximise MG, implying a balanced distribution of load throughout the day. A processing power load analysis is not made in this case, due to the used power model being completely oriented to LTE.

One can see in Table 4.10, the different traffic curves produce different multiplexing gains. Actually, the earlier 2G and 3G technologies may achieve higher gains than LTE. However, as a final result, the immediate conclusion is that the observed gains do not improve as one combines all curves together, the reason being that one is creating more mixed cells when analysing more traffic together (the share of mixed cells increased to almost 32%) while there are fewer cells being classified as office and residential ones. Note that the traffic peak in 4G is at 3 p.m. corresponding to a minimum in 3G one, which is related to radio conditions (frequency and interference dependent) and services provided in each technology, which is out of the scope of this work. Besides that, 4G has the biggest share of traffic

in the city, so it is expected to have a larger contribution to the result. One should have in mind that the share of mixed cells is not the only factor contributing to the result, but also the specific curve of the cell traffic being combined and the absolute difference of traffic in each one of them. The previous explanation is used for justifying the value obtained in 3G traffic, even with a slightly more share of mixed cells. Note that the share of cells classified as office RRHs decreases significantly in the last case (33% instead of 43%), resulting in a lower MG as expected from [ChHC14].

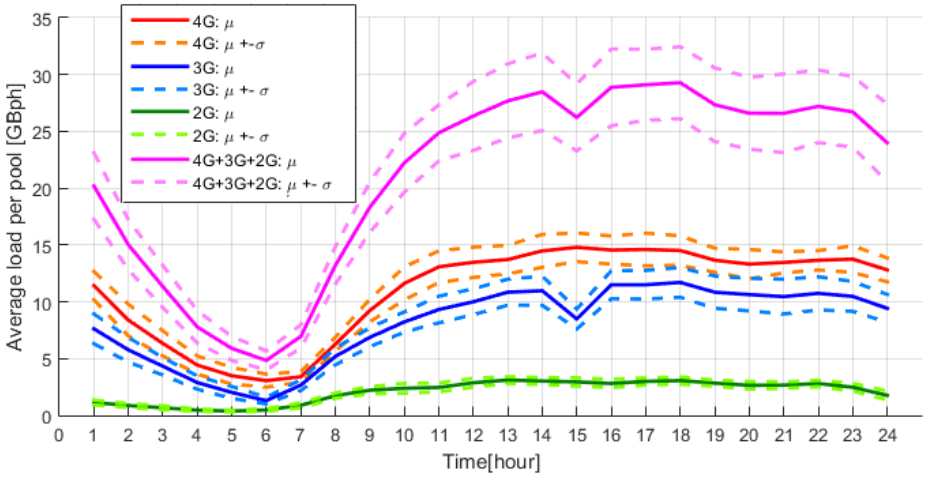

Figure 4.15. Average load per pool with different traffic profiles.

Table 4.10.Multiplexing gain comparison (Maximise MG algorithm).

Traffic Analysed	Multiplexing Gain	Cell Classification Ratios [%]		
		Office	Residential	Mixed
4G traffic	1.31	43.2	37.1	19.7
3G traffic	1.40	46.1	35.5	20.5
2G traffic	1.62	51.2	30.3	18.5
All traffic	1.23	32.7	35.7	31.6

Finally, one can see in Figure 4.16 the average traffic capacity that would be required in order to deal with the maximum traffic demand.

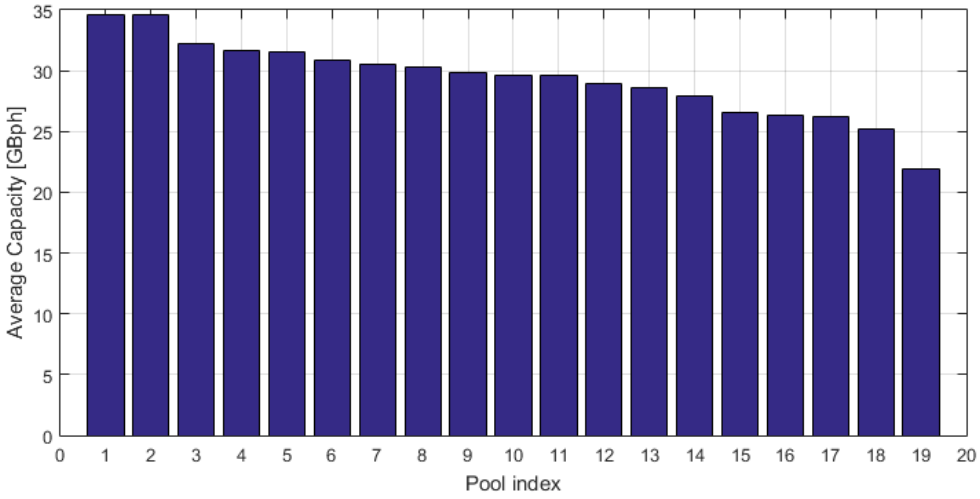

Figure 4.16. Average capacity in each pool.

The first thing to notice is that the balancing of load per pool is not so precise as in the LTE traffic case in the Maximise MG algorithm. One is doubling the average capacity of each pool and also its standard deviation, as expected, as the traffic per hour is also increasing about two times. For the same reasons already pointed out, the value in processing capacity is not represented, because it would fall off of the validity of the used power model (specific for LTE). If one tried to consider all traffic as LTE's, assuming that the current BS can handle the extra one, the model would not have a big impact, because the load is not the most relevant factor in processing capacity.

4.6 Number of Pools Analysis

A C-RAN theoretical deployment would require a single physical location to concentrate all processing functions, decreasing costs and achieve higher gains in multiplexing and cooperative transmission techniques. However, as in the case of the reference scenario, the mobile operators' point of view has another important concern regarding load balancing and network optimisations, besides the aforementioned delay limitations. To study the effect of the number of pools, one presents multiple scenarios in Annex B.

Analysing latency, Figure 4.17 shows its evolution with the number of selected pools (the Minimise Delay algorithm is used).

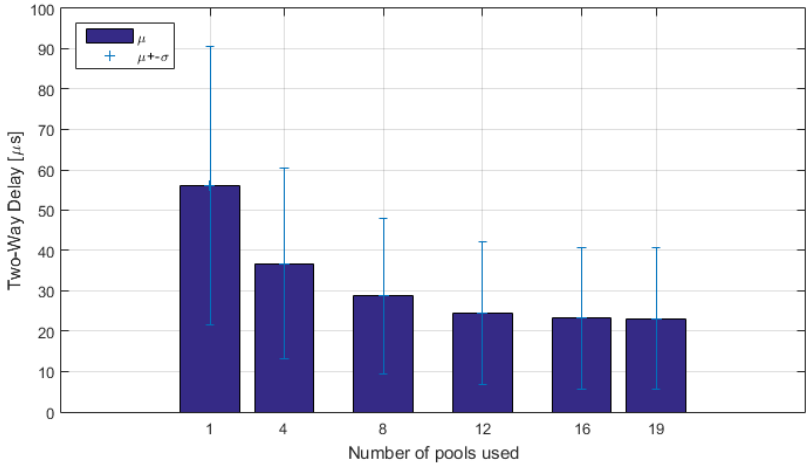

Figure 4.17. Average two-way delay with the number of pools used.

As expected, one gets lower average and standard deviation values as one is using more pools in the scenario. Naturally, the chosen active pools in each scenario were based on the coverage they offer to the network. Although the values in a single pool scenario are still within a certain margin of the theoretical maximum (60 μs), 24 RRHs are already not centralised. One should have in mind that a reduction of 5 times in fronthaul delay (as expected for 5G) would certainly impact on the number of cells served in all studied scenarios. Besides all that, one can see that the difference between having the 19 pools scenario and the 12 pool one is minimal, due to the presence of some pools in close

geographical zones.

This last conclusion is also supported by the capacity handled by each pool. Figure 4.18 and Figure 4.19 show the capacity results obtained with the Load Balancing algorithm. One can see that the use of 19 distributed pools decreases the capacity per pool in 94%, corresponding to a processing power decrease of the same order. The difference between the 16-pool and 19-pool scenario is about 16%.

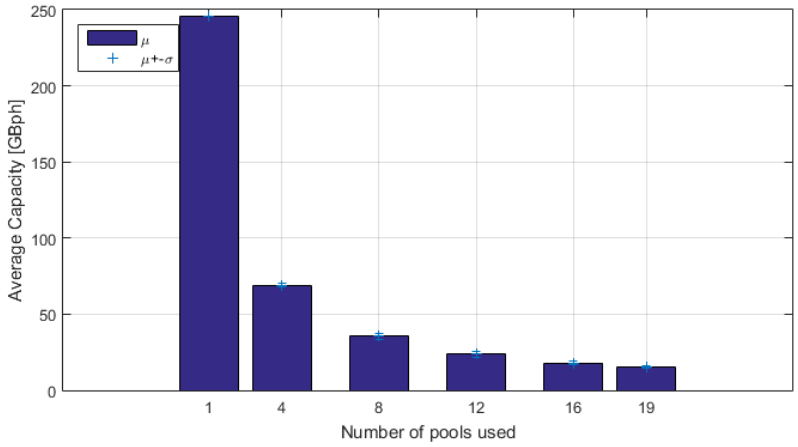

Figure 4.18. Average capacity per pool in GBph with different scenarios.

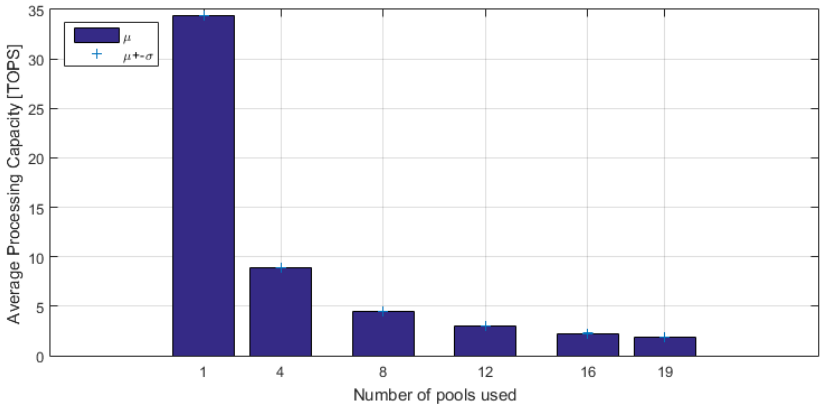

Figure 4.19. Average capacity per pool in TOPS with different scenarios.

Concerning the multiplexing gain involved in this kind of deployment, when running the Maximise MG algorithm, one obtains very close values for both multiplexing and processing of 1.31 and 1.03, respectively. The reason for this fact has already been addressed, being justified by the same traffic curves having to be combined and for the dominance of the much more loaded in the overall gain. Naturally, the gain with under-provisioning is also the same when all RRHs are centralised. With that being said, it is natural that the CAPEX represented in Figure 4.20 appears as a constant. As the model of capital expenditures does not scale with the number of pools, one gets the same values in C-RAN deployments due to very close MGs. But, the same conclusion does not apply for OPEX. The MG is obviously still the same and the savings in maintenance and energy do not differ for the same reason. However, the price of renting increases when more pools are used. Although in the first two cases there are still some RRHs not centralised contributing with higher energy consumption costs, the increase is

almost due to renting and it approximately linear, growing about 46 k€ per each extra pool used.

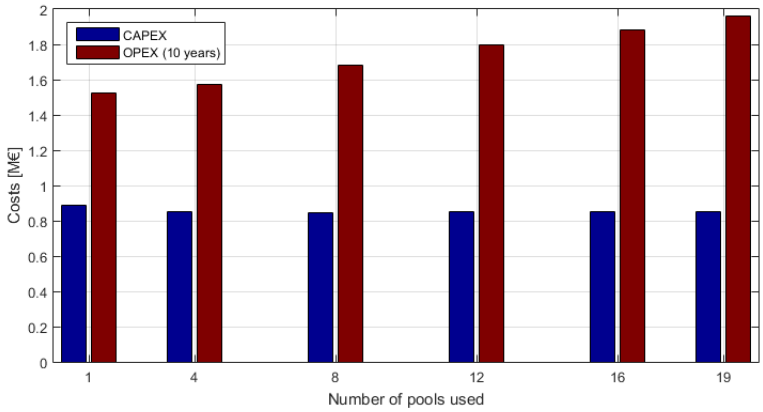

Figure 4.20.TCO with the number of pools in each scenario.

4.7 Daytime Balancing Analysis

As traffic in not homogenous throughout the day, another approach that a mobile operator can elect to implement in its network is the scaling of pools during the day. The main goal of this strategy is to set a capacity limit for each pool as a way of balancing the load, or as a consequence of physical limitations, such as equipment capacity, physical area for hardware accommodation, energy supply, or related problems. To simplify the analysis, one selects three time instances representative of three time periods: 6 a.m. for early morning low traffic, 3 p.m. for labour time peak traffic, and 10 p.m. for residential time peak one. Figure 4.21 shows the overall network load during the day, being clear that 6 a.m. is the lower loaded hour, and that 3 p.m. is the most loaded one, with a 21.7 GBph difference to the reference hour.

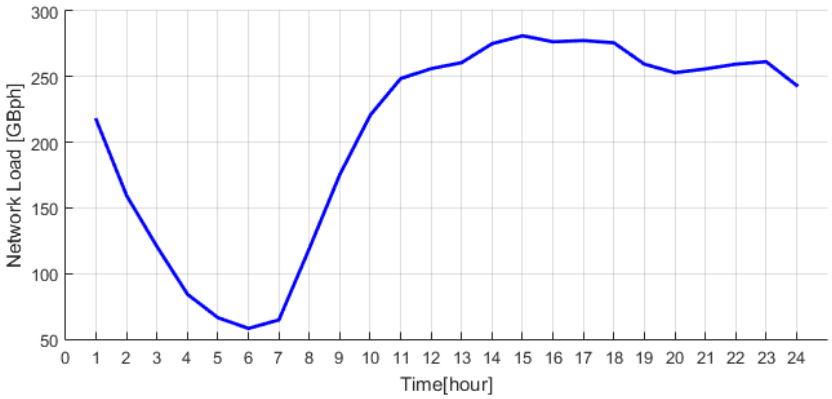

Figure 4.21.Overall network load during the day.

The time instances are used separately in the Minimise Number of Pools algorithm to get the number of pools required during each hour. The used algorithm is presented in Section 3.3.5 with just one time instance as input, results being shown in Figure 4.22. For the capacity limits, one has used the average capacity obtained in the reference scenario, to ensure that all cells are centralised and simulated trough

different values up to 125 GBph. Nevertheless, it is important to state that the capacity unit used is not natural as a restrictive entity; it is used instead of processing capacities and net data rates, as it is the real input analysed in this thesis.

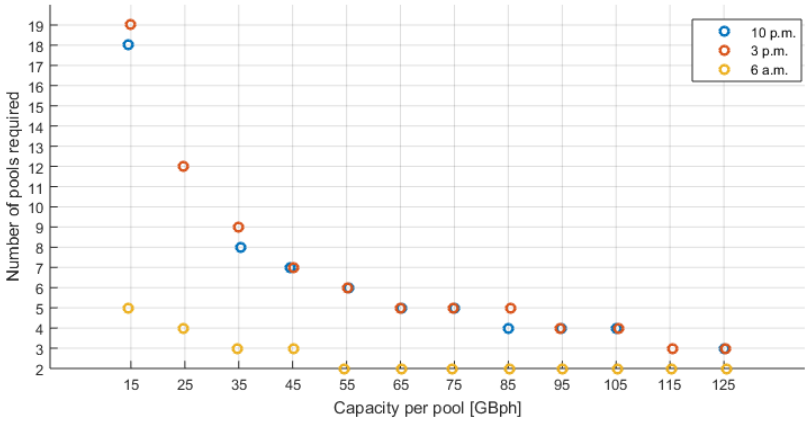

Figure 4.22.Number of pools required with maximum capacity per pool.

The number of pools required during the night is quite below the one during the busiest instances of traffic. In fact, 5 pools are enough to process the 6 a.m. traffic in the reference scenario, but when the values of allowed capacity increase, one can see the difference becoming less relevant. When capacity limits are higher, the number of pools tends to be minimal. For the 6 a.m. case, there is a limit of 2 pools in the scenario, which is actually imposed by distance, as any of the possible pools is able to concentrate all RRHs with a fronthaul constraint of 15 km. Finally, note that, in the largest values of capacity, there is no relevant difference in between pools as only three are enough to process the higher traffic demands. Naturally, this situation represents a waste of resources during the night, as a major part of capacity is unused. Further simulations show that a capacity of 245 GBph is required to support the minimal number of pools (2) at any time instance.

One should have in mind the limitation from the Minimise Pool algorithm. First, the minimal heuristic introduces some losses in multiplexing gain that, combined with variable renting costs, introduce some variation in the projected TCO. Figure 4.23 shows the TCO obtained in multiple simulations running that Minimise the Number of Pools algorithm, now with three combined time instances as input (6 a.m., 3 p.m. and 10 p.m.). One should notice that the cost model does not consider eventual OPEX savings from turning off some pools during night time. In the latter case, there is the possibility to introduce additional savings mostly related with energy consumption. The energy model has not that kind of granularity, and there are also increased costs in more complex switching fronthaul architectures, which is out of the scope of this thesis.

The first peak in OPEX is easily explained by the partial centralisation scenario. As only 90 RRHs are centralised, the cost of renting pools and sites at the same time introduces additional costs to the fully decentralised deployment (0 capacity). Naturally, OPEX tends to decrease, as higher multiplexing gains are achieved. Still, as the minimise algorithm does not take MG and load balancing into account, the maximum MG approach is only achieved by the capacity limit of 55 GBph (6 pools in use). After this point, one can see a flatter CAPEX curve with minimal changes, only due to small MG fluctuations

around that value. OPEX is slowly decreasing due to a reduction of renting until only 3 pools are used to completely serve the area. The values achievable with a capacity of 145 GBph represent cost savings of 48% compared with a decentralised implementation, and 25% compared with reference scenario.

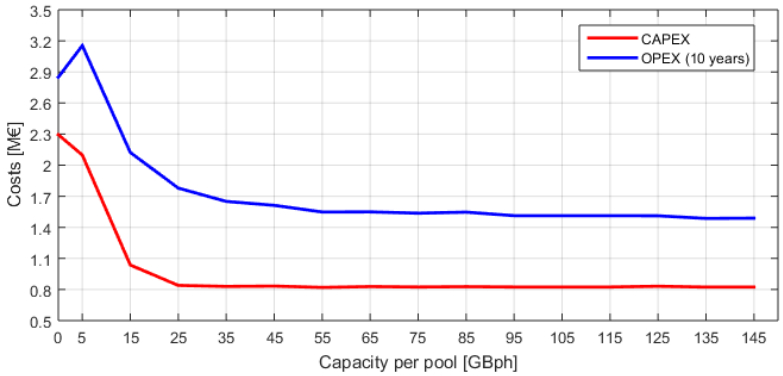

Figure 4.23.TCO of the network with capacity per pool.

4.8 Proliferation Analysis

This section presents the analysis concerning the proliferation layer. It is important to have in mind that traffic estimation is a complex task, as multiple applications and services are being constantly offered and changing traffic demand. In this case, one presents two scenarios that reflect an operator forecast of traffic growth in both pessimistic and optimistic approaches. One also offers an intermediate scenario that works as a more realistic result with an increasing proliferation factor per year, supported by the traffic prediction presented in Figure 1.1 .Traffic evolution is then adapted to a small cell deployment that reflects that growth in a denser network (Section 2.3). Table 4.11 shows the values considered in the analysis. Although one is considering 10 years in OPEX estimation, 5 years of densification are used as in previous work [Maro15]. However, one applies the proliferation factor to the sites themselves, and not to the number of RRHs, resulting in a more conservative number of small cells. Note that 5 years from now, 5G technology is expected to be already tested (2020) and operators' investment may be concerning new technologies. Besides proliferation, one assumes a significant traffic growth in the already deployed macro-cells BSs, considering they are still able to serve higher demands with other technologies, such as CA. This growth is assumed as year 0 for simplicity of the model (growth factor).

Table 4.11. Proliferation factor per year.

Scenario	Growth Factor [%]	Proliferation Factor [%]				
		Year 1	Year 2	Year 3	Year 4	Year 5
Pessimistic	0	54				
Realistic	25	54	57	62	68	75
Optimistic	50	75				

Figure 4.24 shows the site number evolution throughout the years. One site represents only a cell in newly deployed RRHs. These values may seem exaggerated at first, but one should have in mind that

small cells can actually be used instead of Wi-Fi hotspots, justifying the large numbers obtained.

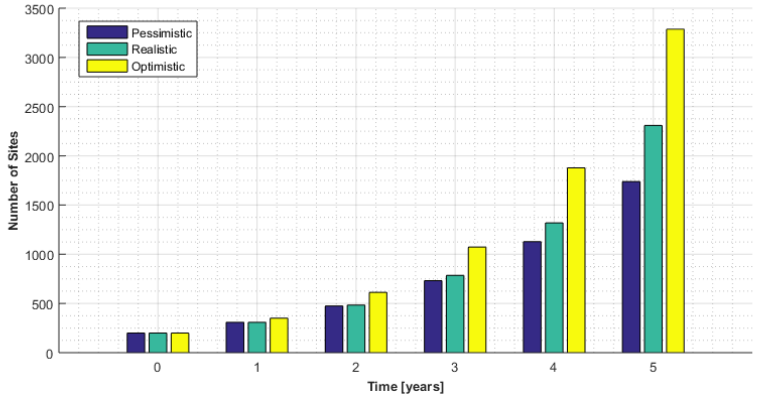

Figure 4.24. Number of sites deployed with the number of proliferation years.

Concerning latency, the results obtained under the Minimise Delay algorithm show a very small decrease in the average value of two-way delay, as expected, as one is mostly deploying new cells in the denser zone of Porto, where there is also the largest number of pools available. The conclusion is that the deployment of new RRHs in denser zones does not increase the delay in the network. However, small cells are often associated with a tight cooperation among themselves and among macro-cell RRHs due to interference, so the values can be lower than 150 μs, but having in mind that at least 114 μs are needed to serve all cells (maximum delay).

Regarding network load, one presents its evolution in Figure 4.25 in the peak hour for the three cases. Proliferation was performed by simply averaging the traffic of the neighbouring RRHs of the new small cell. One can see the difference between the three cases in the two final years. When comparing the overall maximum in five years, there is an increase compared to the reference case in previous sections of about 7.9 for the optimistic case, 4.8 for the realistic and 3 for the pessimistic ones. Actually, the value for the optimistic case is close to the 9-time increase studied in Section 1.1, supporting traffic prediction.

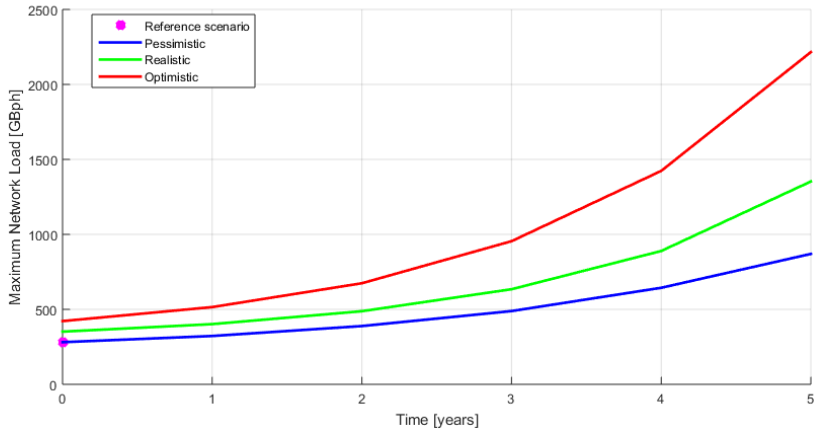

Figure 4.25.Maximum network load with the number of years of proliferation.

The next comment is related to pools capacity. Naturally, with a higher traffic demand, the pools maximum capacity is certainly increasing. As pools are well balanced throughout the scenario

(Section 4.2), one gets the overall network load in Figure 4.25 distributed among 19 pools. Figure 4.26 presents the results, as expected. The final conclusion is that, in 5 years from the reference year, one needs to increase capacity in each pool 7.4-times to handle new traffic demands.

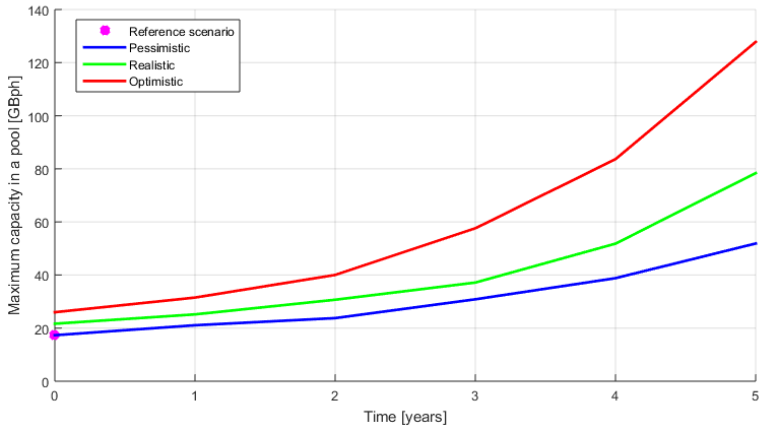

Figure 4.26.Maximum capacity in a pool with the time of proliferation.

Besides the previous analysis, one also needs to analyse GOPS to understand processing capacity demands. Naturally, the evolution of maximum capacity with proliferation years follows the same rational as in traffic per hour. However, one assumes that small cells do not require the same processing power as macro-cells. Processing power scales by changing the value of the computational resolution from 24 to 16 bits, due to the more controlled radio conditions ([DDLo15]). The previous assumptions lead to less processing power required in the improved network. Figure 4.27 shows the results.

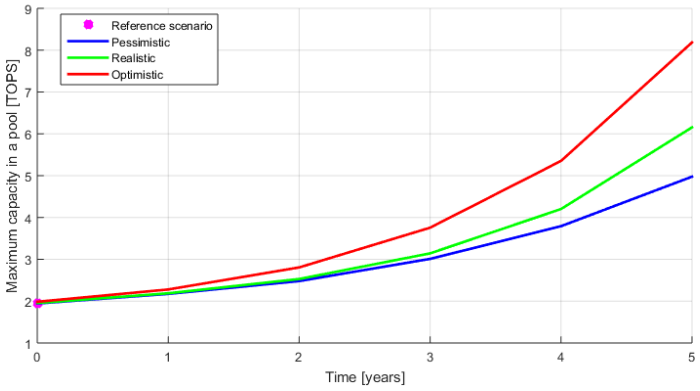

Figure 4.27.Maximum processing capacity in a pool with the time of proliferation.

As small cells do not require the same processing capacity, the growth of the maximum capacity per year is not as high as traffic demand. In fact, one gets an increase of about 4 times (instead of 8) in the optimistic scenario, and about 3 times and 2.5 in the realistic and pessimistic ones, respectively. In year 0, the difference caused by traffic growth in the already deployed RRHs is not much significant, as most of the processing power is not load dependent (see Annex A). The conclusion is that the processing capacity that needs to be installed in the following years of C-RAN deployment is not as high as traffic demand. Still, an operator should be prepared for a significant increase in the investment made in processing boards and interfaces, and in energy and maintenance per year.

To complete the capacity analysis, one uses the capacity limit that led to the minimum number of pools (3 pools with a 115 GBph limit) to analyse the effect of proliferation years. Figure 4.28 shows the results for minimising the number of pools algorithm under the three time samples (6 a.m., 3 p.m. and 10 p.m.).

One can see that the reference scenario of 3 pools is not prepared to handle the traffic growth considered for year 0. In fact, even in the pessimistic scenario, a mobile operator would have to triplicate the number of pools installed in the network. In the optimistic scenario, one can see that the possible pool locations are not enough to handle all traffic. As a reference, a capacity of 125 GBph or an extra pool would be required to ensure that all RRHs are centralised.

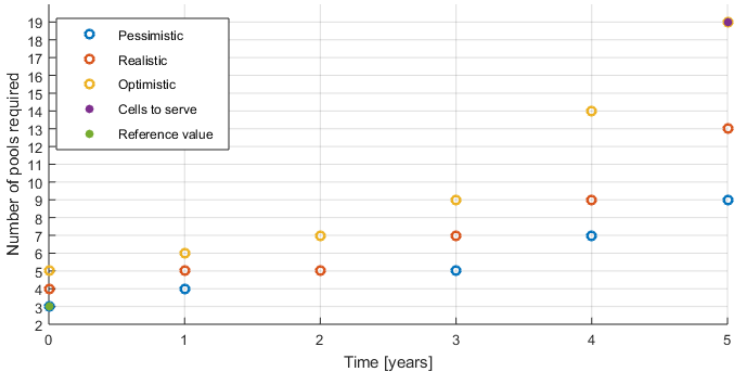

Figure 4.28.Number of pools required with the proliferation time.

The next analysis is important to understand the importance of the centralisation of small cells in the multiplexing gain. Although one has been using an average of neighbouring traffic for the newly deployed RRHs, one has proposed more deployment models of RRHs that represent other possible realistic scenarios (Section 3.3.2). Table 4.12 shows the results obtained for multiplexing gain under the different traffic proliferation methods. The traffic curves used for the specific traffic profiles method are shown in Annex E, being calculated automatically to represent the most pronounced curves of each type. They are normalised, because they are still scaled with the neighbouring cells to reflect different traffic demands.

Table 4.12.Multiplexing Gains Achieved with different traffic models.

Proliferation Scenario	Traffic Average	Pounded Traffic Average	Specific Traffic Profiles
Pessimistic	1.261	1.313	1.398
Realistic	1.257	1.312	1.408
Optimistic	1.248	1.306	1.402

Two main conclusions are drawn from results. The first one is that the proliferation time and factors have a small influence on gain. Although gains follow the same trend along the scenarios, the variance among the three cases is not very significant, even with the number of RRHs difference, since small cells themselves start contributing for traffic estimation and cells ratios become almost constant. The second conclusion is that the deployment of new centralised RRHs is beneficial only if they are installed in specific environments, such as shopping centres, company's facilities or spaces characterised by nocturnal activities. Note that the main difference of the specific traffic profiles (higher gain) is that one

is using very pronounced traffic curves and not averaging the already existent ones. In the deployments of pounded traffic average, one avoids the deployment of mixed profiles, as it can be seen in Figure 4.29, but it is still averaging the office and residential cells, resulting in an almost constant gain.

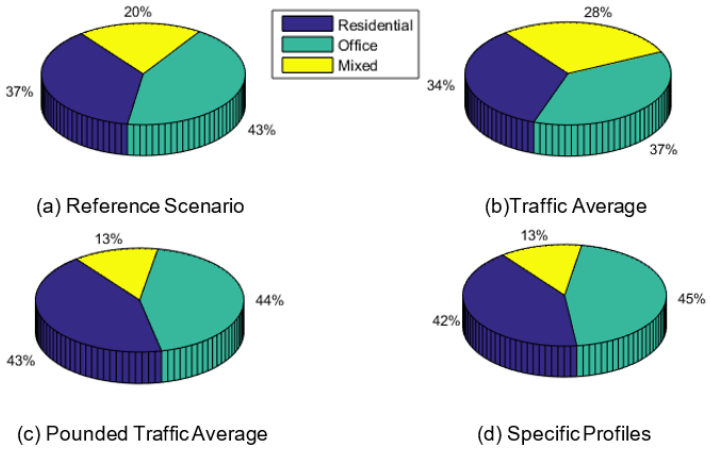

Figure 4.29.Share of cell sites with each classification type.

Note that, with traffic average, one gets more mixed cells, as the combination of residential with office cells results in a mixed cell. The increase in the multiplexing gain is also translated into an increase in gain in processing power. However, one only gets to increase the value of 1.03 to about 1.05 in the 5th year of proliferation in the optimistic case with specific traffic profiles. Regarding the gains with under-provisioning, one gets for the same scenario more loaded cells with less spare capacity (growth factor applied to year 0), causing an overall gain close to the centralisation gain. For a mobile operator, this actually means that the gain of under provisioning is easily lost when the traffic growths and cells start getting their maximum loads. However, this last value also depends on radio conditions (spectral efficiency of the power model), and an operator can still have some profits in cells with very favourable conditions, such as small cells.

A final comment on the MG is the combination of office and residential cells leading to higher gains. Although the ratios of deployment are not easily controllable by an operator (they depend on traffic demand), it still gives an analysis of what kind of cell types should be installed in an area in which traffic offloading s needed. For instance, if a macro-cell is overloaded one can decide if the offloading of traffic should be done in a residential hotspot or in an office centre. Figure 4.30 shows the results with proliferation strategy being the specific traffic profile assignment based on probabilities (Section 3.3.2).

The used normalised traffic profiles are the ones in Annex E, but scaled for the neighbouring traffic demands. Note that one assigns a zero probability for the deployment of mixed cells, as one already proved that they are not good for the gain and they should not be considered to Maximise MG. Results were obtained in the second year of proliferation in the optimistic case under Maximise MG algorithm. If one deploys 60% of residential cells (40% of office ones), the maximum gain in achieved, more than 1.41), this value corresponding to a ratio of 44.4% of global residential cells and 43.8% of office ones, as there were already some deployed RRHs. The conclusion is that a mobile operator should try to keep a balanced deployment of both types of cells. The results come as a consequence of office and

residential cells traffic having comparable peaks (on average).

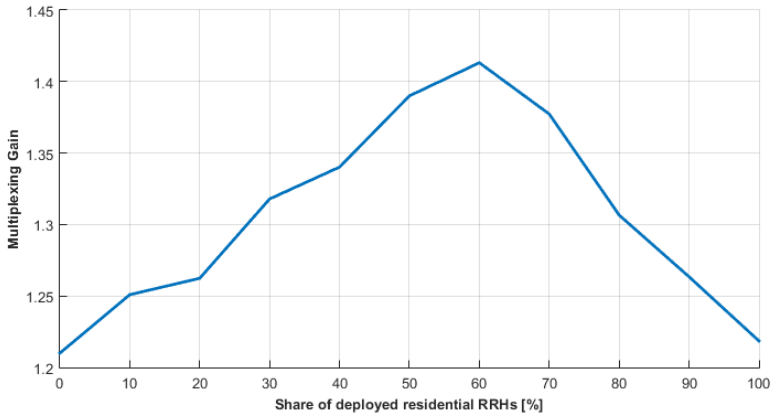

Figure 4.30. Multiplexing Gain variation with shared of deployed RRHs.

The final analysis on proliferation is naturally related with TCO. The rationale used so far needs to be slightly changed, as some of the assumptions on costs are no longer viable. Regarding CAPEX, one should assume that the price of the boards remains constant, as they are supposed to be generic. Regarding hardware, one also keeps the price constant, but having in mind that adding so many RRHs will possibly have an influence on the normalised cost per cell, as extra physical and switching equipment is required. The difference in investment is considered only in CW, as small RRHs are easily installed since they represent smaller equipment. However, there is no significant difference between deploying an entire compact equipment and a small RRH. Regarding OPEX, one considers a smaller BBU consumption, processing power's is about half of macro-cell RRHs'; however, the major difference in energy related to small cells is in the RRH consumption. As maintenance is already depending on CAPEX, one keeps the same assumptions. For the renting component, one should also apply a different model; small cells are small equipment that are not yet using the RRH-BBU split. With that being said, there is no significant difference of deploying a complete base station or a single RRH in C-RAN (no cost savings). It is really important to note that this does not mean that there are no costs in renting for new RRHs. In fact, this component would be the most expensive one for an operator to support and some commercial deals with different partners will be required.

A complete analysis is not presented as the real values of small cells costs in a centralised architecture are not known so far. One of the ideas to retain is that the investment required for the network is proportional to the number of deployed cells. With that being said, one should have in mind that the difference between the pessimistic and optimistic scenarios represents a factor of 2 in the number of cells deployed in the 5th year, which considerably increases the costs related with proliferation. The way traffic grows in the following years has a very relevant impact on the investment required for the operator. Operational expenses are also increasing with the same difference. However, small cells are cheaper in the analogue component (both OPEX and CAPEX) and the digital part does not add so much processing load as macro-cells (less power required). With a very significant number of RRHs to be deployed, renting costs will certainly be the relevant component in TCO. Regarding eventual savings with centralisation, one reinforces that the RRH-BBU split is still not available for small cells. Eventual

advantages come from higher gains that may be achieved with specific deployment strategies. With so many boards and interfaces required for a very high traffic demand, the gain is important to considerably reduce the required hardware. The share of TCO assigned to hardware is not as high as for civil work, but with such values of investment one should consider the centralisation of small RRHs. As for OPEX, the same results of multiplexing gain apply to energy consumption. There are no savings in renting (the main component) and this fact may lead to consider other ways to increase network capacity. Regarding network connections, centralisation implies fronthaul links instead of backhaul ones. Requirements are different, and that will certainly imply higher costs for the centralisation of small cells, which should be compared to the multiplexing gains to define the cost effectiveness of centralised small cells.

4.9 Cost Analysis

This section concerns the TCO model. The first analysis regards the influence of virtualisation with the full share of resources. In fact, one is actually studying the effect of multiplexing gain in cost savings as the remaining parameters are kept constant. Naturally, with no share of resources the virtualisation factor is used as 1 in all pools. One has also added a hypothetical scenario, in which multiplexing gain is 2. Figure 4.31 shows the results.

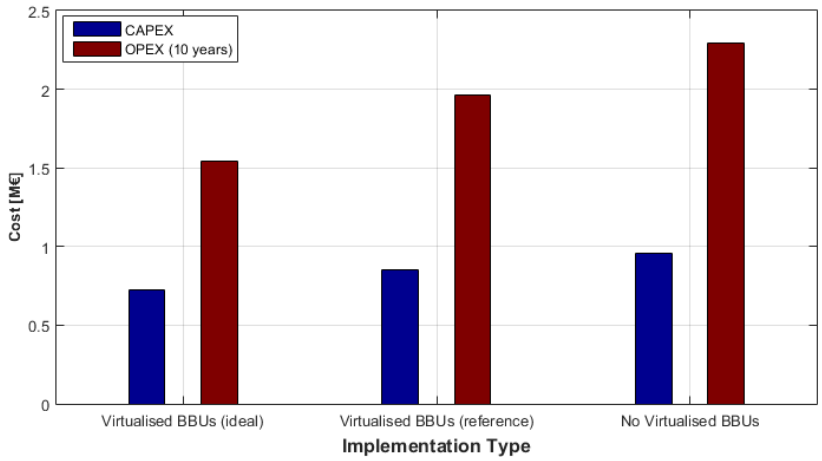

Figure 4.31.TCO comparison with virtualisation in C-RAN.

As expected, results prove that there are additional savings when using a share of resources. When no virtualisation is considered, the savings in CAPEX and OPEX compared to a full decentralised deployment decrease 5% and 12%, respectively. The numbers considering MG are not as high as often treated in literature, but when one considers the high investments that would be required to support new traffic demands, the values obtained represent an important reduction in TCO. The savings are even higher if vendors do not provide sharing solutions to an operator (no benefits for them), as one has already seen that the highest savings are due to civil work reduction. Regarding the ideal scenario, in which the MG is 2, one has obtained an increase in savings of 7% in CAPEX and 14% in OPEX. Note

that the gain may even be higher than 2, if the cells have balanced peaks in more than two time instances. In fact, the maximum value is equal to the number of time instances under analysis. The used value corresponds to a scenario in which half of the cells have the peak at 3 p.m. and the other half at 10 p.m. with the same peak value, which would be a regular and symmetrical movement of people from home to work, i.e., corresponding to a virtualisation factor of 50% (half of the boards required). Following the same analysis, one has used different considerations in gain throughout the work. In order to see the difference that multiple considerations have on savings, one has defined the cost reduction in other scenarios. If the processing gain in DL is used, the savings are only 1% above the no virtualisation scenario in both CAPEX and OPEX. However, one has seen that the processing gain is higher for UL, so in an optimistic scenario, one considers for the gain the value for UL with under-provisioning. The results obtained are an increased value of 4% and 11% relative to the reference scenario.

Another analysis concerns the renting case. The results obtained are relevant for a different business model, in which boards are leased instead of bought. Note that a leasing strategy can also be applied to more hardware components. However, as the values of these components are combined into a single component in the cost model (together with physical structures, licences, etc.), one uses the assumptions that they represent an initial investment. The leasing model assumes that the leasing of a board is 200 €/year, and that the costs assigned to boards stop contributing to maintenance expenses. Figure 4.32 shows the results.

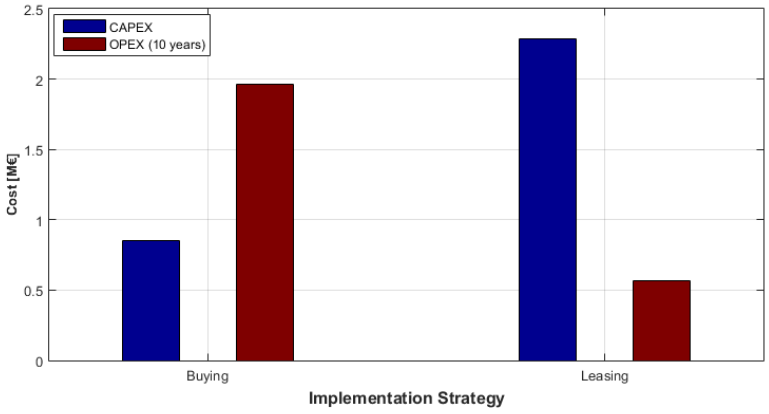

Figure 4.32. TCO with different strategies in C-RAN.

One can actually see a large difference between OPEX and CAPEX. For a mobile operator, the conclusion to have in mind is that CAPEX can be alleviated by 50% while increasing OPEX about 14%. Regarding savings, one has about 71% and 28% of savings in CAPEX and OPEX, respectively, compared with a leasing decentralised deployment.

The next analysis regards hardware savings (civil work not considered). Virtualisation is a concept used in standard computer boards usually known as GPPs. Although one has been using the prices of specific hardware boards, GPP implementations are more scalable and can also have an influence on hardware savings. This analysis focuses on the influence of the GPP factor ρ (introduced in Section 3.2.5). Besides this factor, one has also used a factor ν to account for hardware complexity associated with switching schemes and intra-pool communication. Figure 4.33 shows the effect of these factors. Note

that 1 is the value for both factors in the reference scenario. The reason for this assumption is that realistic values are being used, and the values achievable with later technology are not known. One can clear see that the reduction with multiplexing gain is outweighing the price paid for centralised auxiliary equipment, licences and physical structure (savings of almost 2%). However, it is a well-known fact that GPPs are cheaper than specialised hardware [CCYS14], and even if the latency limitation imposes some restrictions on full GPP implementation, there is still the possibility to decrease the value of σ to less than 1. In this case, there will be savings on hardware cost, due to cheaper equipment combined with multiplexing gain (assuming ν as constant).

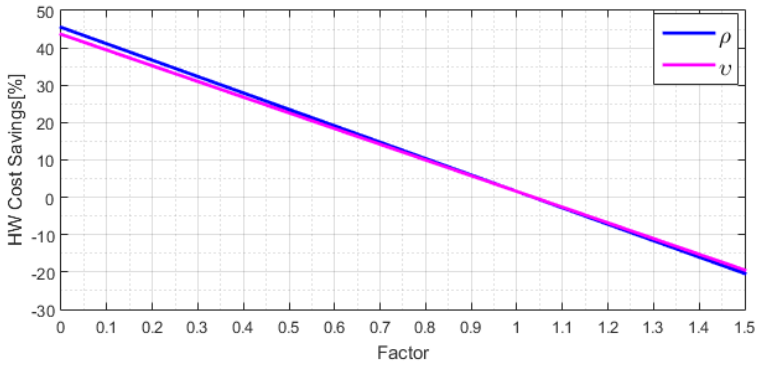

Figure 4.33.Influence of cost factors on HW savings.

Regarding additional complexity in hardware savings, one can see the influence of the ν factor being almost the same as GPP cost reduction. Although C-RAN equipment is not usually treated as an additional cost in literature, the reference scenario assumes a price per cell that is 1.5-times higher than the regular deployment. As this factor accounts for more than electronics, it is not certain that it will decrease in the following years, and may outweigh multiplexing gain.

A final analysis is done in energy consumption. Figure 4.34 illustrates the influence of energy efficiency on OPEX savings (η factor). The first conclusion is that this value is actually important in overall OPEX savings as it can cause significant variations. The second conclusion is supported by the fact that general purpose processors are less efficient than specialised hardware. Depending on technological implementation, the loss of efficiency is translated in less savings. In fact, if GPPs are 1.5-times more consuming than specific hardware, the savings become less than 25% instead of the reference 31%.

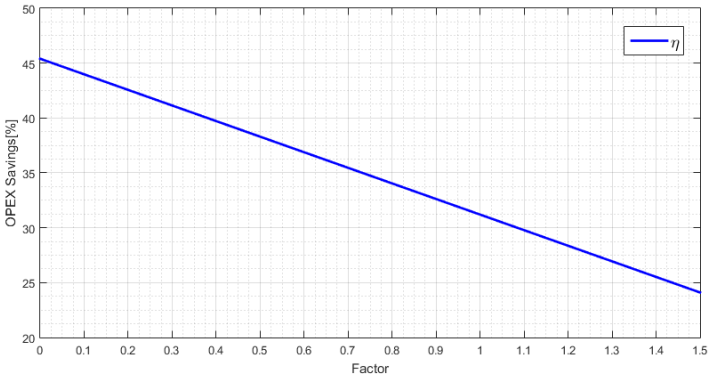

Figure 4.34.Influence of cost factor on OPEX savings.

Chapter 5

Conclusions

This chapter finalises this thesis and presents conclusions. Future work perspectives are also stated in this chapter.

The main goal of this work was to analyse the performance of the deployment of a C-RAN architecture in an already existent LTE-A network. One has developed a model and a computational tool to receive the LTE network as input and to define the C-RAN configuration and export results regarding capacity in the pools (in traffic per hour or processing power), latency, multiplexing gain and costs of investment and operational expenditures to characterise cost savings.

In Chapter 1, a global view of mobile communication systems is presented. A particular emphasis is given to the 4th generation of mobile communications system, LTE-A. The relevance of the subject of the work is also presented and supported by traffic predictions and current technological challenges stimulating new strategies. Then, the general concept of radio access network in the cloud is introduced.

Chapter 2 provides a theoretical support of the technologies being treated in the thesis. The chapter starts with an explanation of LTE-A, which includes the identification and description of the main components of the network architecture, as well as their functionalities and characteristics, and the radio interface description, focused on important aspects such as access schemes and resource blocks concepts. Then, a more detailed description of some of the new concepts in which this thesis is focused is provided. First, C-RAN is contextualised by unfolding two of the most discussed topics in nowadays networks, such as SDN and NFV. After that, the C-RAN concept is discussed in detail. Based on the most important works and definitions in the literature, the new network architecture is presented. The most significant theoretical advantages are listed and justified, as well as the current challenges holding its deployment and requiring further study. Following that, small cells are described to contextualise some of the possible implementation of C-RAN in futuristic scenarios. The chapter finishes with the state of the art stating the most relevant studies found in the literature addressing C-RAN, specifically the performance parameters and the applications scenarios studied in the work.

In Chapter 3, the model is fully described, with flowcharts of the most important algorithms and with detailed information on the parameters under study. The chapter starts with a brief description of the problem being solved. Then, the parameters under study are presented and justified with the relevant formulas and reference values found in the literature. The considered parameters are latency as the propagation time in fronthaul links, load of the cells in GBph, processing capacity in GOPS to measure computational power, MG to characterise peak traffic reduction with centralisation, and TCO to quantify cost savings. After that, the model overview is made in its structural representation and in a more specific way, supported by flowcharts of the algorithms proposed. The model developed in this thesis was made from scratch, except the proliferation layer which was built upon previous work. Note that the starting point of the implementation is the information provided by a mobile terminal.

Regarding the model, one has conceptually divided it into three main components. The first one is the proliferation layer, developed to add the possibility to study futuristic implementations of denser and more heterogeneous networks with small cells deployment. This layer can be adjusted for different time periods and for different proliferation factors representing different growth rates of the networks; it is built upon previous work [Maro15] and adds the traffic proliferation component and some structural changes. The second layer is the assignment based on performance parameters. The idea behind this layer is to decide to which pool should an RRH be connected to under different metrics. The used metrics are

related to the performance parameters and consist of multiple heuristic algorithms to obtain the best overall result of each one of them. First, the minimal delay algorithm attempts to connect each RRH to the closest pool in the scenario. Then, one proposes a Load Balancing algorithm working in two iterations. The first iteration connects the RRHs that only have one possibility to the corresponding BBU. The second one connects each RRH to the less loaded pool. Besides this one, one presents the Maximise MG algorithm working in a way similar to the load balancing one. After assigning the RRHs with only one option of connection, the algorithm tries to connect each RRH based on their traffic profile, attempting to combine traffic curves that produce a flat profile of daily traffic. Finally, a Minimise the Number of Pools algorithm is also proposed based on a known problem in mathematics and a proposed heuristic to account for all the time instances during the day with different loads. The idea is to account for multiple peaks of traffic during the day and assign RRHs to the pools respecting the capacity limits in every time instance under study. The third layer consists of the cost model. One has decided to study only the digital components of the network for comparison of local and centralised implementation, as analogue components are the same in both cases. Therefore, one has constructed a model for CAPEX mostly based on licences, hardware costs (equipment, boards and interfaces) and estimated civil work costs. For OPEX, one has constructed a model based on energy consumption of digital components, maintenance of hardware equipment and renting of sites or pools. The final note is that fronthaul costs are not being assumed as the operator already owns a deployed fibre network. However, CPRI requirements may still need additional investment.

Chapter 4 starts by defining the scenario of Porto and multiple assumptions taken as reference. The reference scenario is then analysed, which was also used for algorithm comparison. One has used Minimise Delay, Load Balancing at 10 p.m. and Maximise MG algorithms to compare the results of the different metrics. Results prove that the three heuristic proposed produce the expected results. When the selected strategy is to minimise the distance in an RRH-BBU connection, one gets considerable less delay in the first algorithm, as more than half of the RRHs are at 2 km or less away from the respective pool. In the other two approaches, one gets a substantial number of RRHs (about 15%) at a distance beyond 12 km from their pools. However, in the last two strategies, one is naturally aiming at different goals. The main conclusion drawn is that both algorithms achieved similar results in both metrics (Load Balancing and MG). As future work, it is interesting to consider optimisation techniques to Maximise MG with more powerful mathematical tools to consider all the possible combinations that are not being tested in the iterative heuristic proposed.

As a final result, one got as best values an average capacity of 15.2 GBph under the Load Balancing algorithm and traffic gain of 1.31 under the Maximise MG one. As expected, the capacity value is caused by a balanced distribution of the traffic through the 19 pools of the scenario. The same happens when one analyses the processing capacity, achieving an average capacity of 1.8 TOPS. Note that the processing capacity is several orders of magnitude under the results in [Maro15] and [WeGP13]. The reason for this fact is that the authors of those works were considering a previous power model that was not oriented to the real number of operations of each physical function, but for the energy consumption based on a conversion factor. In what concerns MG, the results have the same magnitude of previous studies that pointed out to a maximum of 1.6. However, all the studies from the authors in [ChHC14] do

not consider the presence of mixed cells that do not have a pronounced traffic profile and that do not improve the MG. It is also important to note that previous studies were based on two curves only, assigned randomly through the scenario and not providing such a realistic environment. Still, one has also treated the gains using GOPS as a unit. The results prove that the gain of processing capacity are quite under the theoretical ones from previous analysis with traffic (only 1.03 as gain). The main reason justifying the latter result is that most of the current computational power in LTE is derived from non-load dependent physical functions. As the load is the factor producing MG, the overall value of processing gain is only about 10% of the multiplexing gain. In what concerns technology, this interpretation can support the idea of changing the splitting point and implement some L1 functions in the RRH. This change would be beneficial to under loading the CPRI requirements, but would also complicate the cooperative transmission techniques in that layer. However, a significant number of cells are not working in their maximum load or not processing in the worst radio conditions or service requirements (spectral efficiency of the model). With that being said, there is also a gain achieved due to the under provision of the RRHs processing capacity. Note that this gain is not due to centralisation in C-RAN, but it is still related, as a smart centralised architecture can be easily scaled to address more load if necessary by adding more processing boards. The aforementioned gain obtained was around 1.28 and confirms the experimental results in [BCJK12]. This value gets close to the fully load-dependent results of MG.

In what concerns cost reduction, one got savings for the investment of about 63%. The value is close to savings obtained in [AHGr15] when the authors assume a similar virtualisation factor, but twice the one presented in [CMRI11]. The reason for such a difference is that the authors only consider the costs in civil work and materials as 24% of the overall CAPEX. Naturally, the study was done in other continent and there will certainly be some differences in the business model. In this work, civil work represents about 71% of costs and the main reason of savings. It is important to note that electronics is becoming cheaper every year. Besides that, one should note that the comparison done in this work is focused only on the centralisation and not in the change of technology from legacy BS to RRH-BBU schemes, which certainly justifies the not so relevant savings in equipment (AC is not considered, for instance). The same fact justifies the not so pronounce savings obtained in energy. In fact, the legacy macro-cell BS requires a specific AC equipment that is no longer needed with RRHs. The savings of 31% are still far away from the suggested 50%, but it is highly dependent on the renting models of the operator as it is the main source of cost reduction. A detailed study analysis of fronthaul investment and maintenance should be interesting to add in future work. Future work should also be focused on building upon the cost model with more detailed hardware considerations and to make it scale with the number of cells per pool to compare costs with different algorithms.

The analysis of performance parameters dependence with maximum fronthaul distance was also done to study the effect of much more stringent values of delay allowed by the technology. Although most of the literature considers 150 μs as a limit, one should have in mind that the fibre links under study are not straight lines, that GPPs in the pool would have a negative impact on the overall latency of LTE, and that 5G technologies are aiming at 1 to 5-times lower latency. The results show that there is a significant impact on the number of RRHs that can be centralised with the possible pools. Naturally, the partial

centralisation has an impact on MG and TCO. One has seen that the MG reaches its top value at a limitation of 7 km. As a consequence, the results in CAPEX and OPEX do not suffer a significant difference with restrictions higher than that value, as most of the load is already centralised. In more stringent values, one can see the influence of MG and mostly the civil work costs. The final conclusion is that the 15 km restriction can be lowered down to 7 km without significant loss in costs, but that under that value the partial centralisation can have a significant impact in TCO.

One has also analysed the scenario using UL traffic curves. The average DL traffic per hour in a pool is about 8-times higher than UL, as a result of asymmetric services, such as video streaming and web browsing. However, one could actually see that this result is not translated into more computational power, as UL requires more processing load due to UL radio conditions. Besides, one has also pointed the fact that processing power in UL is more load-dependent, as there is a significant variation during the day. This fact is also supported by the higher gain in processing power (around 1.23) and by higher gains with under provisioning as most of the cells are processing less than required in that direction. This result proves that UL should be considered when analysing gains from the current technology viewpoint. In a fully load-dependent perspective, one has seen that UL traffic profiles are more pronounce, but that the contribution to an overall gain is not significant due to higher DL traffic. In future work, it would be interesting to combine both analyses. More serious considerations regarding hardware could also be added to properly characterise technologies implementing more load dependent solutions.

Then, one has presented an analysis of traffic profiles. The idea is to understand the impact that the traffic being handled by legacy technologies has on LTE. This also has in mind the possibility of multi-RAT support, although the current cost model is not prepared for this kind of convergence. An integration of 2G and 3G in the cost model is interesting for related and future work. In terms of load, LTE is handling almost the same data as the other two technologies combined. The convergence of traffic would mean doubling the required capacity in the network. Besides, the effects on multiplexing gain are not so optimistic as expected by the results of legacy technologies. Although 2G and 3G traffic curves present more pronounce traffic profiles and higher gains, the fact is that the combination of the three is producing flatter profiles (more mixed cells). In fact, 3G and 4G traffic types are not matching, and it is certainly related to radio conditions and services used that are out of the scope of this work.

Another analysis was focused on the daily load variation. The goal was to add another interesting perspective for an operator consisting of minimising the number of pools. Note that a deployment like this can offer a much higher level of CoMP as it maximises centralisation. Besides, with a futuristic level of switching, one can change the BBU-RRH connections during the day. With that being said, one adds a capacity limit in the simulator to apply the Minimise the Number of Pools algorithm. In this analysis, one has seen a significant difference between the traffic at 6 a.m. and 3 p.m. In an opposite way, one could see that the peak of the network at 3 p.m. is very close to values found for 10 p.m. One has also concluded that at least two pools are required to serve the entire scenario, even with infinite capacity. However, one considers 115 GBph as a reasonable limit that requires 3 pools in the network for the following analyses. In this scenario, an operator can choose to switch off one of the pools to serve the entire traffic during the night. A cost study was also done to see the influence of the number of pools in

TCO. With capacity under 15 GBph, one gets the same conclusions of partial centralisation as in the latency analysis. For higher values than 15 GBph, all the load is handled by the centralised pools and there are only small variations in CAPEX due to MG variations. Regarding OPEX, one can see the difference caused by the required number of pools. With 125 GBph limit, one actually saves 24% in OPEX, compared with the reference scenario with centralisation. However, one should note that adding complex switching schemes will increase the costs that should be studied in future work.

A proliferation study based on small cells is also presented. One has proposed three scenarios based on an operator viewpoint reflecting optimistic, pessimistic and realistic results in terms of traffic. The results of traffic growth of the optimistic case get close to the prediction presented in Chapter 1 (9-times increase in 5 years). This growth is supported by a significant increase of sites in the network. In fact, one is adding almost 3 000 small RRHs throughout the scenario in 5 years. Naturally, the results show an increase in the traffic capacity of each pool of the same order of the traffic growth. As 19 pools are available throughout the scenario, the algorithm of Load Balancing proves to be an effective solution to distribute the load among all pools. However, the growth in processing power does not increase as much as the traffic per hour, as the processing power required for a small cell is lower than the macro cells. Besides, the growth factor applied to year 0 scales the traffic per hour, but only increases the load-dependent processing complexity functions. For the same reason, the under-provisioning gain is something that decreases when cells start getting their full load. However, note that there is still some gain due to cells with more favourable radio conditions (possible on specific types of small cells deployment) or less consuming type of services on average that causes less power load in the pool that should be confirmed in future work with realistic values. Still regarding MG, the results show that an operator should avoid deploying small cells in areas that do not have a specific traffic profile (mixed). The main idea is to install small cells in spots where the traffic profile is very specific and known for an operator. The continuous deployment throughout the years with the same strategy will result in an almost constant gain. One has also proved that the ratios of deployment should force a balanced number of each type of cells as the peak traffic in office and residential cells are very close. This result is slightly different from the one used in [HCAC15], as the authors are considering that office cells have higher peaks than the residential one. Finally, a qualitative analysis on cost savings was done to possible integrate small cells in the cost model. As this kind of equipment is not yet available with the RRH-BBU split, it is important to use future realistic values to confirm the analysis with numbers. If correctly deployed, one can achieve higher gains with small cells and increase the savings in those components. However, civil work difference will not have the same impact on the price per cell and savings will not be significant. The same conclusion was drawn for renting, as there is no meaningful difference in physical dimension of small cells and small RRHs. The results obtained for MG should be compared to the fronthaul cost of deploying fronthaul links instead of backhaul links, to define if there are significant advantages in centralising small cells besides the obvious CoMP gains, or if a mixed architecture in two layers (C-RAN with standalone small cells) should be considered, as suggested in [ACHC15].

Finally, a cost analysis is also presented to study some important aspects on TCO. The first conclusion is that even in a non-virtualisation scenario, in which there is no share of physical resources, there are significant cost savings mostly due to civil work reduction. However, the values are 5% (CAPEX) and

12% (OPEX) under the reference scenario. The reduction in savings is significant in the case of a high investment that will certainly be needed to support the new traffic demand. In an optimistic scenario, one has considered a multiplexing gain of 2 in all pools. The savings increased 7% and 14% in CAPEX and OPEX, respectively. Although the scenario represents an idealistic case, the values are important for understanding how much savings MG can introduce in an operator's TCO. Then, one has made some changes in the business model of the operator to see how TCO would change if the computer boards were leased and not bought. The results have shown that the initial investment can be reduced by half, but still having to pay more in each year expenditures. The business models depend on the contracts that an operator uses with suppliers. A final analysis was done to see the influence of some technological dependent factors, such as cost reduction of GPPs, energy efficiency and extra complexity in centralised pools. The results prove that the price of the boards and the added complexity of C-RAN introduce very similar variations in the hardware savings. One has seen that the cost of GPPs will certainly be lower than the reference scenario, but the same thing cannot be said about added complexity. Regarding energy efficiency, the value of this factor is expected to increase with IT processors as they are less specific and efficient. As future work, it is interesting to see how different configurations (hardware accelerators, full IT data centres, etc.) affect TCO and virtualisation.

As additional suggestions, the work should be complemented with a more in-depth study of radio conditions. Realistic data would not only introduce more accurate results in the processing capacity as it would make possible to characterise CoMP gains. The study of CoMP is also a promising technique to use with small cells. Besides all of these works, a final advantage of C-RAN is related with offloading of traffic in data centres, so a study of C-RAN network beyond the baseband data is also interesting.

Annex A

Processing Power Reference Scenario

Reference values of processing capacity for the calculation of the required processing power per BBU are presented.

As introduced in Section 3.2.3, the calculation of the processing power is made through the reference scenario that is characterised as:

- Bandwidth: 20 MHz;
- Single antenna and single stream;
- Modulation: 64-QAM;
- No channel coding;
- Load of 100%;
- Quantisation of 24 bits.

The reference values for the identified physical layer operations and for the other layers are shown in Table A.1 and Table A.2, respectively. The value used depends on the study being done (DL or UL).

Table A.1. Processing power for physical layer (adapted from [DDLo15]).

Operation name	DL [GOPS]	UL [GOPS]
Predistortion	10.7	00
Filtering	6.7	6.70
Up/Down sampling	2	2
TD non-ideal. estimation/compensation	1.3	6.7
FFT/IFFT, FD non-ideal.	4	4
MIMO precoding	11.3	0
Synchronisation	0	2
Channel estimation and interpretation	0	3.3
Equaliser compensation	0	3.3
Equalisation	0	2
OFDM Modulation/Demodulation	1.3	2.7
Mapping/Demapping	1.3	2.7
Channel Coding	1.3	8

Table A.2. Processing power for high layer functions (adapted from [DDLo15]).

Operation name	DL [GOPS]	UL [GOPS]
Control	2.7	1
Network	8	5.3

Annex B

BBU Pool Positioning

The positioning and identification of the possible pools location is presented.

This annex presents the BBU pool locations for multiple scenarios.

Figure B.1. Pools identification (reference scenario).

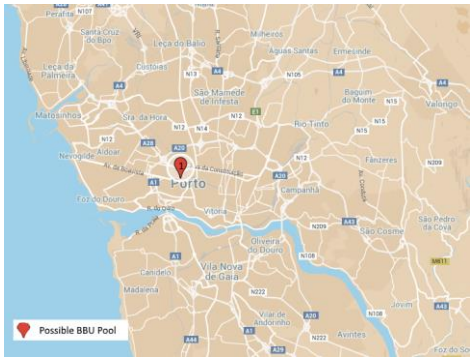

Figure B.2. Pools identification (scenario with 1 pool).

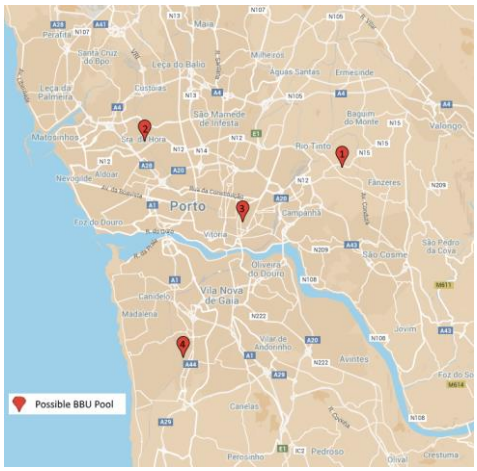

Figure B.3.Pools identification (scenario with 4 pools).

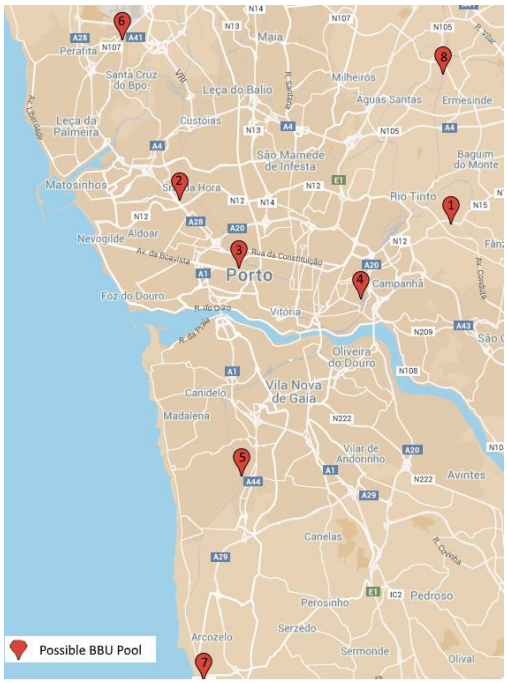

Figure B.4.Pools identification (scenario with 8 pools).

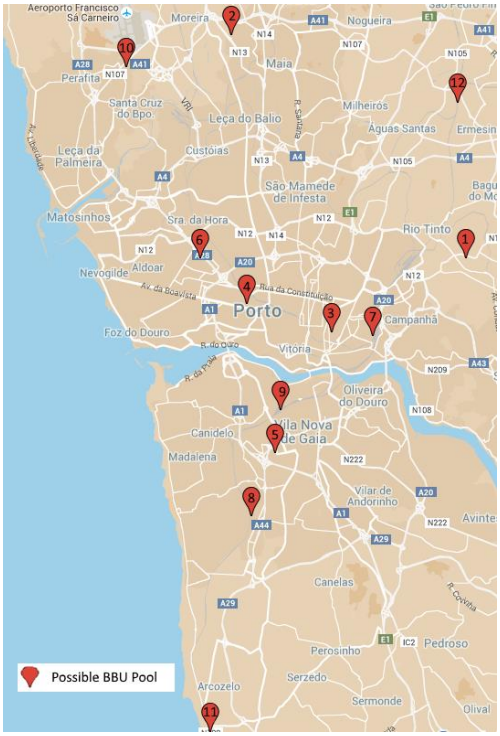

Figure B.5.Pools identification (scenario with 12 pools).

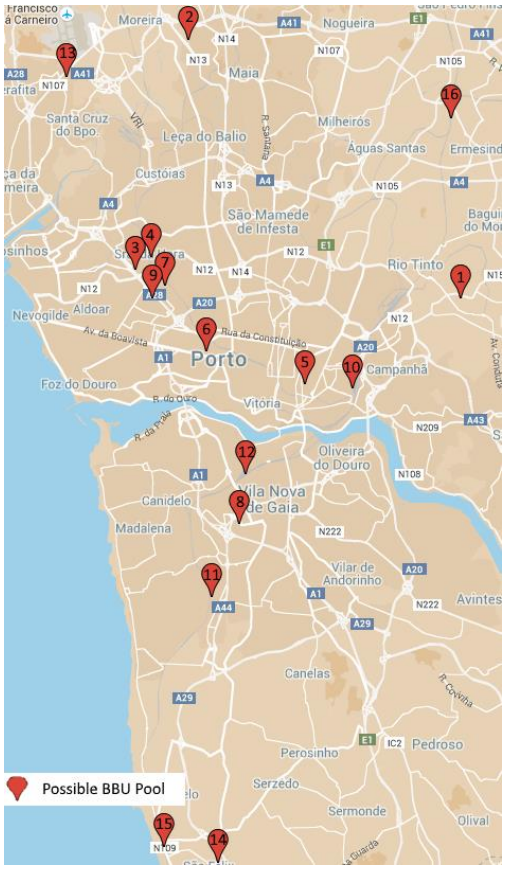

Figure B.6.Pools identification (scenario with 16 pools).

Annex C

Classification of cells

Annex C

Description of the algorithm used to classify cells as office, residential or mixed.

In order to classify the cells based on their traffic profiles, one uses the metric in Figure C.1.

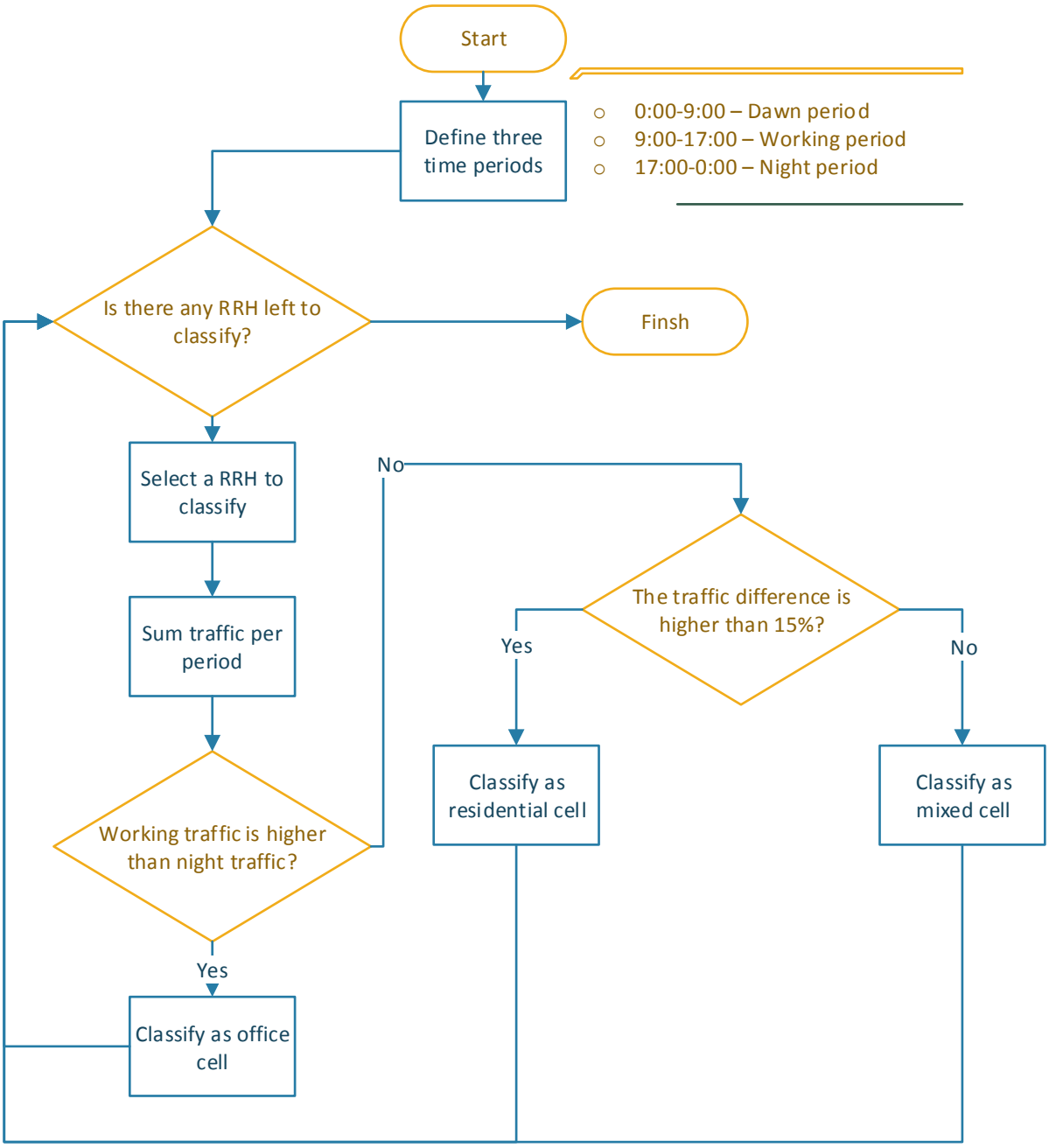

Figure C.1. Algorithm for cell classification according to traffic.

The algorithm classifies the cells according to a comparison between the overall traffic in two time periods, working time and night time. One also adds a threshold to identify a cell that is neither office or residential, as it happens in the scenario. This option is taken to actually study the influence of sites that to do not presented peaks of traffic, as these are the ones that decrease the overall gain.

Annex D

Average Traffic Profiles

Annex D

Average traffic profiles for DL and UL for each type of cells.

Figure D.1 and Figure D.2 present the average per cell of the LTE traffic variation along the day in DL and UL, respectively.

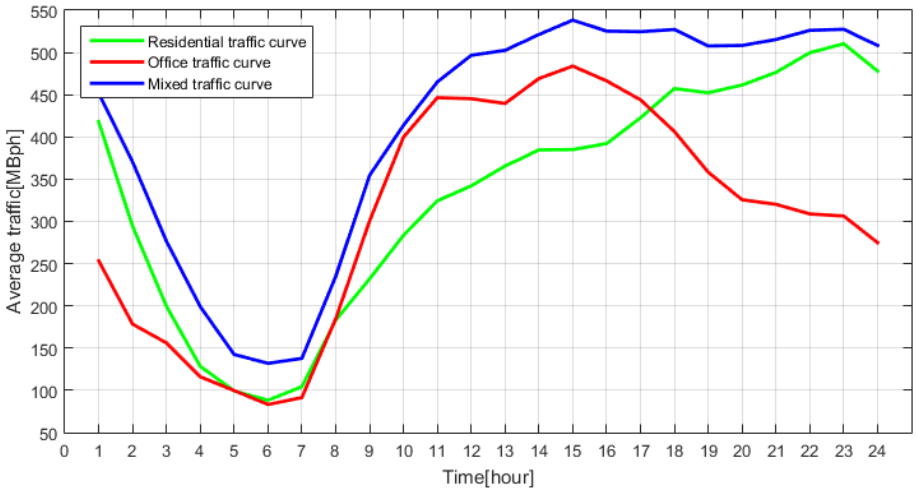

Figure D.1. Average DL traffic variation in LTE (per cell).

Figure D.2. Average UL traffic variation in LTE (per cell).

Figure D.3 and Figure D.4 present the average per cell of the 3G and 2G traffic variation along the day, respectively. Figure D.5 presents the combined traffic of all technologies.

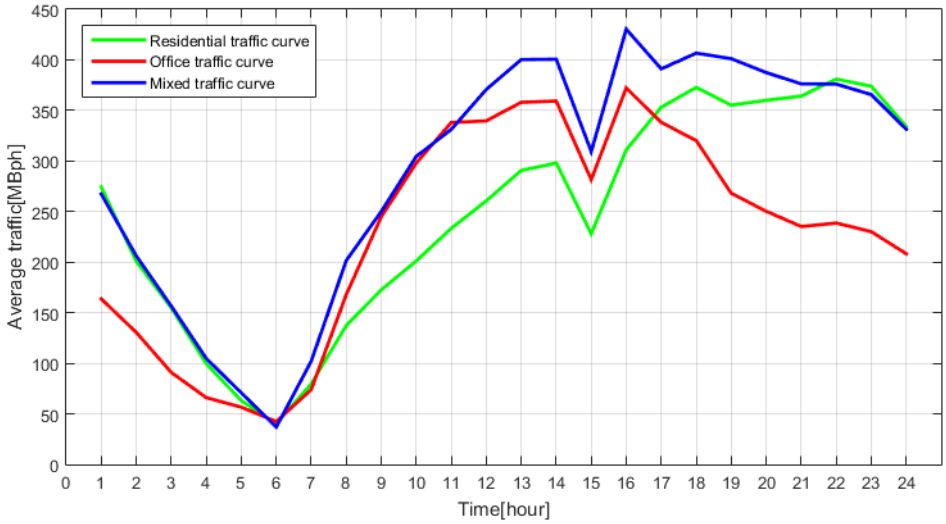

Figure D.3.Average DL 3G traffic variation (per cell).

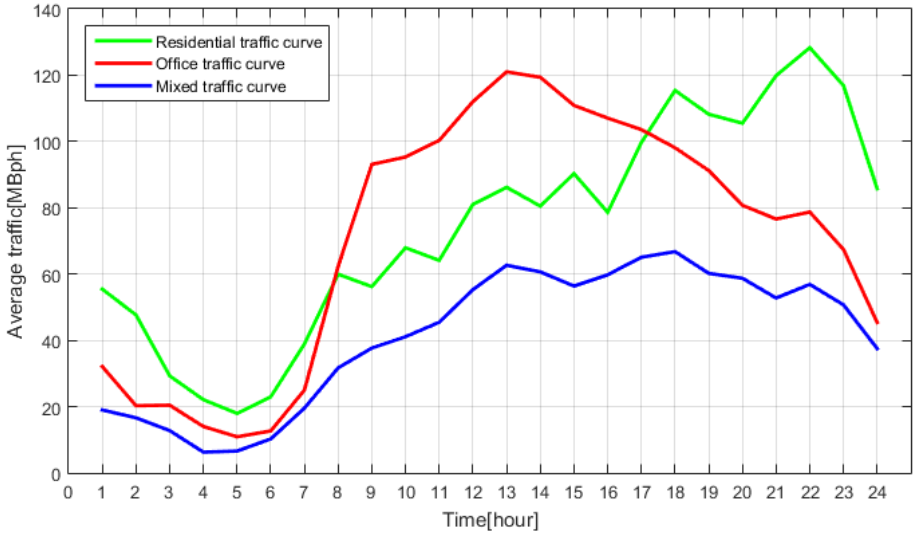

Figure D.4.Average DL 2G traffic variation (per cell).

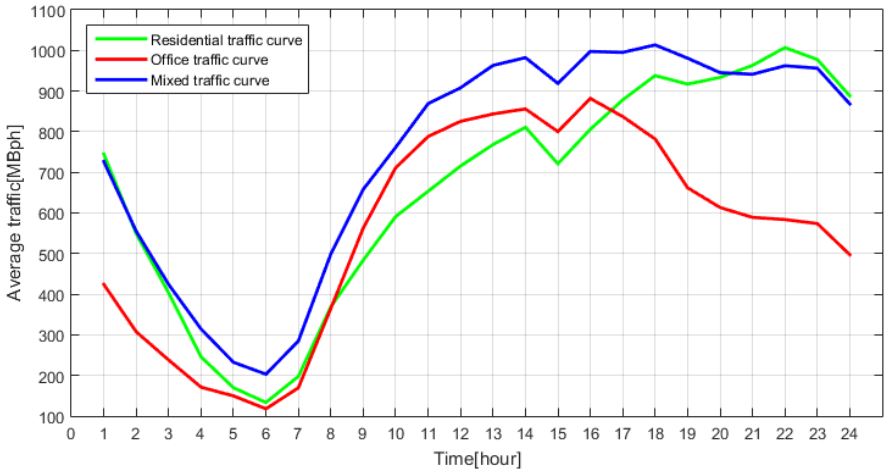

Figure D.5.Average combined DL traffic variation (per cell).

Annex E

Standard Traffic Profiles

Traffic profiles of each type of cell used in proliferation.

Figure E.1 shows the curves selected by the simulator to work as traffic profiles for proliferation analysis.

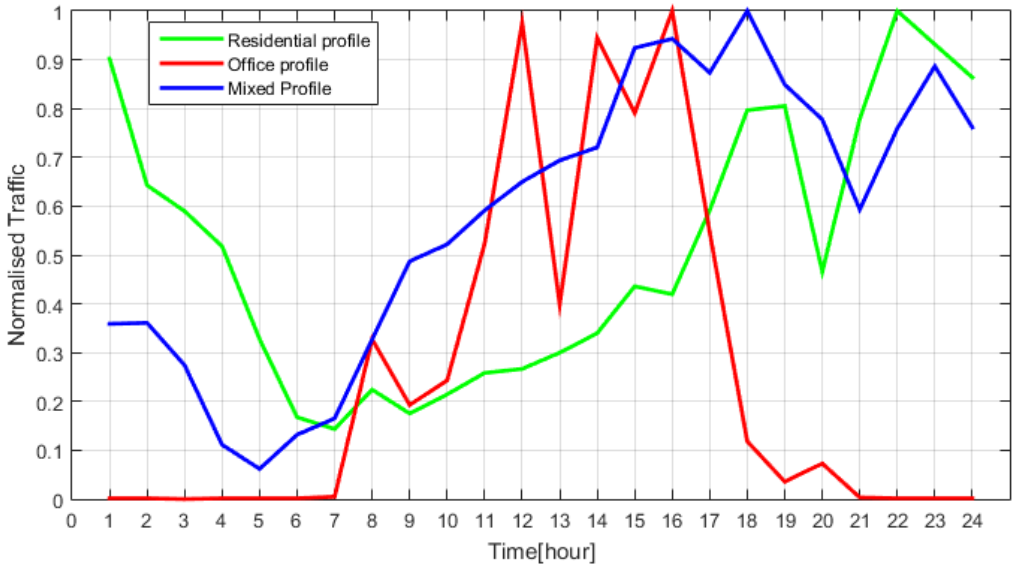

Figure E.1.Normalised traffic profile used for each type of cell.

Annex F

Proliferation Scenarios

RRH coordinates and classification of each proliferation scenario.

Figure F.1, Figure F.2, Figure F.3 and Figure F.4 represent the proliferation scenarios used in the thesis. The traffic classification presented is based on averaging the neighbouring RRHs traffic.

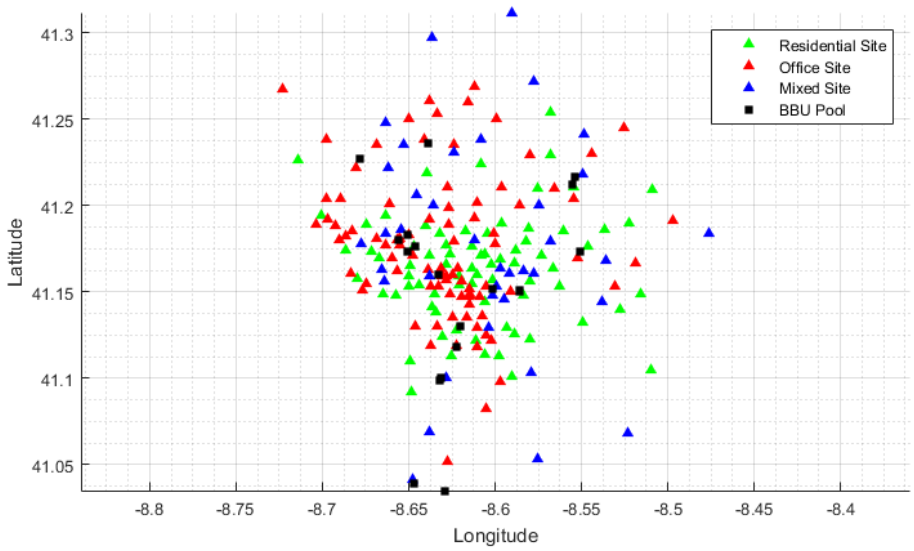

Figure F.1. Reference Scenario.

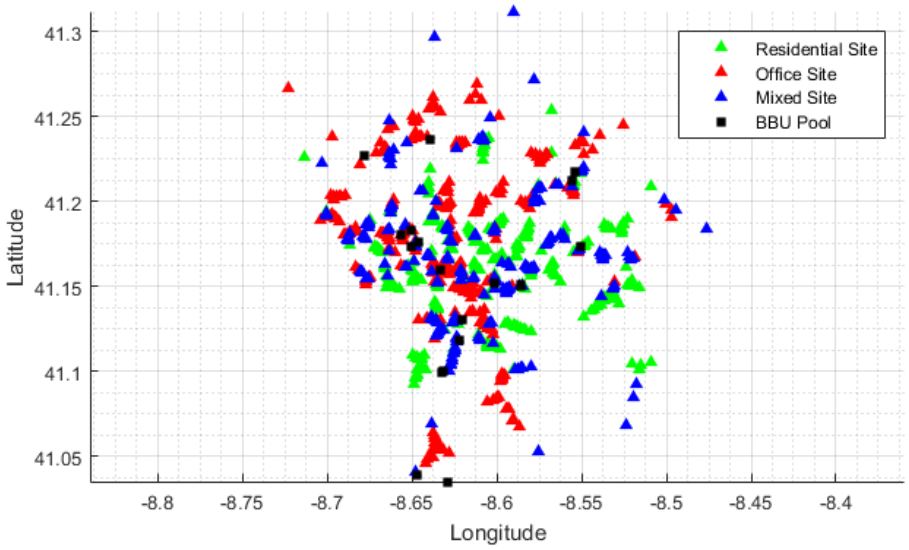

Figure F.2.Pessimistic Proliferation Scenario (traffic average strategy).

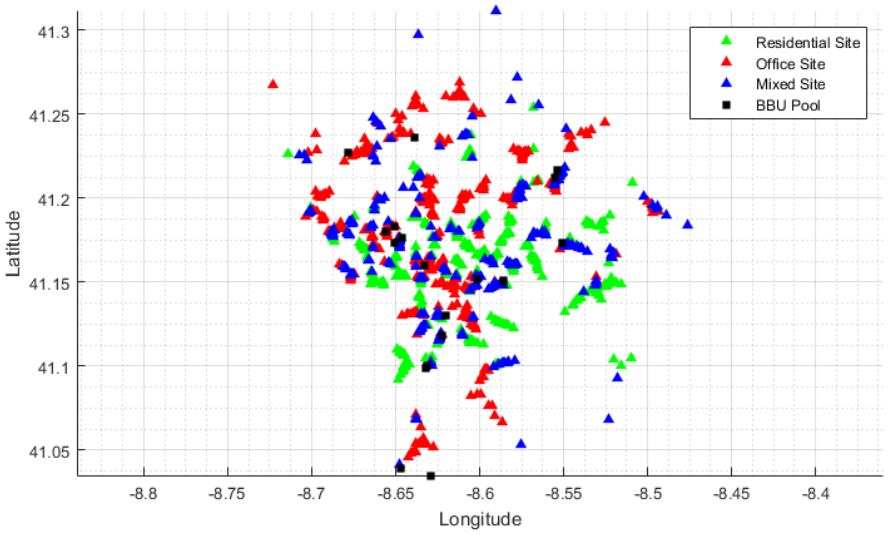

Figure F.3.Realistic Proliferation Scenario (traffic average strategy).

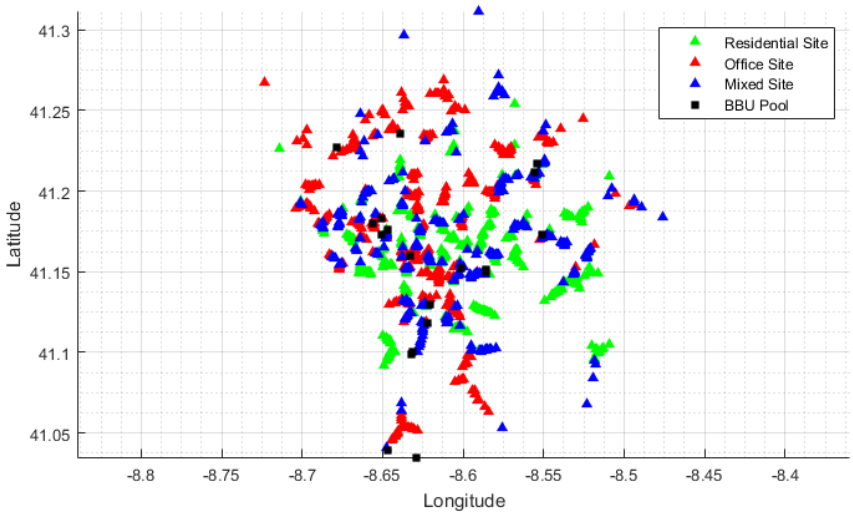

Figure F.4.Optimistic Proliferation Scenario (traffic average strategy).

Annex G

User’s Manual

This annex explains how to configure and run the simulator.

G.1 Reference Software

The program was developed entirely in Matlab's environment. All functions are distributed through scripts files that can be run from the software. As a reference, the used version was 2015a with the most common Matlab's toolboxes.

The available software was used to implement auxiliary functions. These files were obtained in MathWorks's File Exchange website and are identified. The downloaded scripts are organised together with the new ones.

Although the file was mostly used for a single scenario, the simulator is designed to run for other scenarios if the input files are correctly configured.

It is also possible to change or to add work to the current scripts with new functionalities to add new features to the simulator. The functions are commented and some adjustable parameters not editable by the inputs are identified. Additional information on each script is written in the "README.txt" file.

G.2 Run the Simulation

To run the simulation, one should copy all the scripts to the default directory of Matlab or configure other directories. In the software, one should be able to run the simulator by running the script "main.m".

In order to work properly, one should have the following files:

- "Configurations.txt", the main document of configurations. In this file one can change the directories name and input files name.
- RRHs positioning file in .csv format. The name of the file is set in the "Configurations.txt". The file consists only in the coordinates (longitude and latitude) of the RRHs in the scenario. It is edited with a simple text editor.
- Possible pools positioning in .csv format. The name is also set in the configuration file. The format is the same as the RRHs positioning file.
- Traffic information file in .xlsx format. As the previous ones, this file should also be set in the configuration document. Note that is a mandatory document if traffic information is required for the simulator. The format required for the input table is explained in G.3.

G.3 Simulator Configuration

To run the simulator from the start, one should adapt the “Configurations.txt” file in order to adapt it to the scenario under study. The configurations file should be inserted in “C://Users/Data” path as default. The way to configure the file is explained in the comments above each line in the default document. Besides the input files name and location path, the following inputs are mandatory:

- Virtualisation flag. Boolean value used to indicate if traffic values should be accounted or not.
- Study flag: Boolean value to indicate what kind of unit is used for traffic (GOPS and MBph are available).
- Maximum radius: maximum distance between RRHs and pools.
- Neighbourhood radius: used for density functions in prediction (reference value).
- Criteria of assignment: name of the algorithm to be run as an assignment strategy.
- Maximum capacity: numeric limit for traffic in pool (according to the study flag).
- Slots per cluster: how many cells can be processed together in the same cluster as connected boards (size of a shelf).
- Clusters per pool: how many pools can a data centre have in its physical location (size of a rack).
- Year of proliferations: how many years should be accounted in the proliferation layer.
- Years factors: array consisting of the proliferation factors for each year (must have the same length as the number of years).
- Traffic prediction type: name of the traffic prediction strategy in new cells created in proliferation layer.
- Number of hours to be considered for traffic strategies.
- Hours to be considered for strategies: time instances to consider for load balancing and multiplexing gain approaches (use the 24h hours of the day for more realistic results).
- Direction of analysis: DL or UL traffic to consider.
- Base station type: Number of different base station that should be configured in the scenario. The different base stations are deployed with the distance to the reference point.
- Reference point coordinates: latitude and longitude of a reference location in the scenario (centre of the area is advised) useful for multiple auxiliary functions.
- Type of traffic to study: Flag to indicate the traffic that should be studied. Note that to properly use this option, the input file with traffic information must contain entries for more than one technology.
- TCO strategy: Flag to indicate if the strategy is ‘buying’ for acquire processor boards or ‘renting’ for a leasing solution.

The input file with traffic information should contain a site ID column, a total cell column with the number of cells in that site, a column identifying the hour of the traffic information and two columns with the traffic (DL and UL) for each technology. Note that the file should have data compatible with the flag indicating the technology under study. Figure G.1 shows a real example of one site with 2G, 3G and 4G data.

	A	B	C	D	E	F	G	H	I
1	ID	TOTAL CELLS	HOUR	Traffic_4G_DL (MB)	Traffic_4G_UL (MB)	Traffic_Dados_3G_DL (MB)	Traffic_Dados_3G_UL (MB)	Traffic_Dados_2G_DL	Traffic_Dados_2G_UL
2	1	3	0	169,13	113,08	601,90	326,87	144,38	28,85
3	1	3	1	56,17	108,34	201,33	326,72	24,03	9,52
4	1	3	2	68,22	109,64	270,28	169,63	35,34	8,83
5	1	3	3	78,21	118,92	202,24	140,15	113,46	17,81
6	1	3	4	73,43	144,90	1 161,37	160,28	23,38	9,49
7	1	3	5	90,99	122,54	1 407,22	193,81	118,35	22,70
8	1	3	6	111,61	134,37	491,30	47,68	92,35	28,54
9	1	3	7	252,28	220,45	484,01	132,99	125,98	25,16
10	1	3	8	655,74	259,10	844,91	283,65	220,76	45,79
11	1	3	9	840,14	280,25	885,93	279,26	262,76	57,64
12	1	3	10	630,33	245,55	969,95	214,13	330,58	67,20
13	1	3	11	630,92	258,36	827,81	248,20	444,99	74,34
14	1	3	12	342,57	271,72	934,61	341,19	643,98	119,06
15	1	3	13	593,09	420,86	1 040,42	200,38	366,15	128,63
16	1	3	14	789,41	294,76	852,19	201,00	369,44	87,29
17	1	3	15	977,87	266,10	695,43	194,43	388,79	83,80
18	1	3	16	890,40	287,22	1 218,62	297,73	399,52	86,96
19	1	3	17	820,57	283,22	962,06	171,28	254,46	61,24
20	1	3	18	538,89	278,07	1 213,17	204,06	349,54	65,44
21	1	3	19	455,20	290,49	717,84	181,51	312,00	77,56
22	1	3	20	334,77	196,06	872,08	182,49	196,41	43,71
23	1	3	21	272,11	119,81	954,18	144,94	300,88	50,33
24	1	3	22	441,77	120,43	505,81	147,17	374,34	52,97
25	1	3	23	346,54	119,14	655,73	116,43	334,84	64,74

Figure G.1. Example of input file configuration with traffic information.

G.4 Output Files explanation

The entry describes the outputs of the program. Although multiple plots are generated in the source code of the file, one considers these as auxiliary graphs and not as the main outputs. However, all the auxiliary files and code available for results are also available. The outputs files are:

- “clusters_output.txt”: Clusters configuration with the RRHs assigned to each pool.
- “clusters_capacity.txt”: Overall traffic in each one of the pools in the 24 hours of the day.
- “data_output.txt”: Multiple results of the simulation run such as costs, distance, multiplexing gain, load, percentage of cells centralised and neighbourhood.
- “unserved_output.txt”: List of cells that were not possible to serve and the reason for it (capacity or distance).
- KML files with the RRHs and BBU pools locations that can be directly imported in Google Maps.

Some of the plots active in the simulation are the locations of RRHs and their classification according to traffic profiles, the average traffic curves for both UL and DL, the required capacity for each pool and the costs graphs.

The data output file contains information regarding the following metrics:

- Summary: main input parameters identification.
- Distance and delay: average and maximum results in kilometres and microseconds.
- Network configuration: share of centralised cells, share of used pools and peak traffic in the network.
- Load balancing: Results of mean and standard deviation of traffic per pool in each time instance

and in the overall capacity.

- Daytime balancing: information of the gains achieved (according to input file metric, GOPS or GBph) and the share of cells classified as office, residential and office.
- Costs output: information of the CAPEX and OPEX values in decentralised and centralised strategies.
- Program info: auxiliary data such as run time of the simulation.

References

[3GPP13a] 3GPP, Technical Specification Group Radio Access Network, *Scenarios and requirements for small cell enhancements for E-UTRA and E-UTRAN (Release 12)*, Report TR 36.932, V12.1.0, Mar. 2013.

[3GPP13b] 3GPP, Technical Specification Group Radio Access Network, *Small Cell enhancements for E-UTRA and E-UTRAN – Physical layer aspects (Release 12)*, Report TR 36.872, V12.1.0, Dec. 2013.

[3GPP14] 3GPP, Technical Specification Group Radio Access Network, *Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN) (Release 12)*, Report TR 25.912, V12.0.0, Sep. 2014.

[3GPP15] 3GPP, Technical Specification Group Service & Systems Aspects, *Policy and Charging Control Architecture (Release 13)*, Report TS 23.203, V13.5.1, Sep. 2015.

[ACHC15] R. Al-obaidi, A. Checko, H. Holm, and H. Christiansen, "Optimizing Cloud-RAN Deployments in Real-life Scenarios Using Microwave Radio", in *Proc. of EuCNC'15 – 24th European Conference on Networks and Communications*, Paris, France, June 2015.

[ADFJ09] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, and S. Parkvall, "LTE: The Evolution of Mobile Broadband", *IEEE Communications Magazine*, Vol.41, No.4, Apr. 2009, pp. 44-51.

[AHGr15] M. Andrade, A. Hamidian, K. Grobe, "Cost Models for Baseband Unit (BBU) Hotelling: from Local to Cloud", *in Proc. of ICCN 2015 – 4th IEEE International Conference on Cloud Networking (CloudNet)*, Niagara Falls, ON, Canada, Oct. 2015.

[ANAC11] ANACOM, *Information on multi-band spectrum auction (3)*, Public Consultation, Lisbon, Portugal, Dec. 2011 [Online]. Available: http://www.anacom.pt/render.jsp?contentId=1106646&languageId=1#.VfvukDZRGCh [Accessed: Sep. 19th, 2015].

[ASBD15] A. Alyafawai, E. Schiller, T. Braun, D. Dimitrova, A. Gomes, and N. Nikaein, "Critical Issues of Centralized and Cloudified LTE-FDD Radio Access Networks", in *Proc. of ICC 2015 – IEEE International Conference on Communications*, London, United Kingdom, June 2015.

[ASRa15] M. Arslan, K. Sundaresan, and S. Rangarajan, "Software-Defined Networking in Cellular Radio Access Networks: Potential and Challenges", *IEEE Communications Magazine*, Vol.53, No.1, Jan. 2015, pp.150-156.

[AvCI15] A. Popovska, H. Christiansen, and V. Iversen, "Cell Deployment Optimization for Cloud Radio Access Networks using Teletraffic Theory", in *Proc. of AICT 2015 – 11th Advanced*

International Conference on Telecommunications, Brussels, Belgium, June 2015.

[BCJK12] S. Bhaumik, S. Chandrabose, M. Jataprolu, G. Kumar, A. Muralidhar, P. Polakos, V. Srinivasan, and T. Woo, "CloudIQ: a Framework for Processing Base Stations in a Data Center", in *Proc. of MOBICOM – 18th Annual International Conference on Mobile Computing and Networking*, Istanbul, Turkey, Aug. 2012.

[CCYS14] A. Checko, H. Christiansen, Y. Yang, L. Scolari, G. Kardaras, M. Berger, and L. Dittmann, "Cloud RAN for Mobile Networks – A Technology Overview", *IEEE Communications Surveys & Tutorials*, Vol.7, No.1, Sep. 2014, pp.405-426.

[ChCB13] A. Checko, H. Christiansen, and M. Berger, "Evaluation of energy and cost savings in mobile Cloud Ran", in *Proc. of OPNETWORK 2013,* Washington, D.C., USA, Aug.2014.

[ChHC14] A. Checko, H. Holm, and H. Christiansen, "Optimizing small cell deployment by the use of C-RANs", in *Proc. of European Wireless 2014 – 20th European Wireless Conference*, Barcelona, Spain, May 2014.

[CMRI11] China Mobile Research Institute, *C-RAN The Road Towards Green RAN"*, White Paper, Beijing, China, Oct. 2011.

[CJLS14] L. Chen, H. Jin, J. Seo, and V. Leung, "An Energy Efficient Implementation of C-RAN in HetNet", in *Proc. of VTC Fall 2014 – 80th IEEE Vehicular Technology Conference*, Vancouver, BC, Canada, Sep. 2014.

[Corr15] L. Correia, *Mobile Communications Systems*, Lecture Notes, Instituto Superior Técnico, Lisbon, Portugal, 2015.

[DaPS11] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband (1st edition)*, Academic Press, Oxford, United Kingdom, 2011.

[DDLo15] B. Debaillie, C. Desset, and F. Louagie, "Flexible and Future-Proof Power Model for Cellular Base Stations", in *Proc. of VTC Spring 2015 – 81st IEEE Vehicular Technology Conference*, Glasgow, United Kingdom, May 2015.

[Eric15] Ericsson, *Ericsson Mobility Report*, Public Consultation, June 2015 [Online]. Available: http://www.ericsson.com/mobility-report [Accessed Oct. 16th, 2015].

[ETSI12] ETSI, *Network Functions Virtualisation: An Introduction, Benefits, Enables, Challenges & Call for Action*, White Paper, 2012 [Online]. Available: https://portal.etsi.org/NFV/NFV_White_Paper.pdf [Accessed: Sep. 7th, 2015].

[ETSI13] ETSI, "Network Functions Virtualisation (NFV); Architectural Framework", GS NFV 001 (v 1.1.1), Oct. 2010.

[FPHG14] L.S. Ferreira, D. Pichon, A. Hatefi, A. Gomes, D. Dimitrova, T. Braun, G- Karagiannis, M. Karimzadeh, M. Branco, and L.M. Correia, "An Architecture to offer Cloud-Based Radio Access Network as a Service", in *Proc. of EuCNC'14 – 23rd European Conference on Networks and Communications*, Bologna, Italy, June 2014.

[Fuji14] Fujitsu Network Communications, *The Benefits of Cloud-RAN Architecture in Mobile Network Expansion*, White Paper, 2014 [Online]. Available: http://www.fujitsu.com/us/Images/CloudRANwp.pdf [Accessed: Oct. 10th 2015].

[GeRK12] C. Gessner, A. Roessler, and M. Kottkamp, *UMTS Long Term Evolution (LTE) - Technology Introduction*, Rohde & Schwarz, Appl. Note 1MA111, July 2012 [Online]. Available: http://www.rohde-schwarz.de/de/service_support/downloads/application_notes/ [Accessed: Sep. 21st, 2015].

[GLDP12] Z. Guebretensaé, K. Laraqui, S. Dahlfort, F. Ponzini, L. Giorgi, S. Stracca, J. Chen, Y. Li, J.Hansryd, and A. Pratt, "Transmission Solutions and Architectures for Heterogeneous Networks Built as C-RANs", in *Proc. of CHINACOM 2012 – 7th ISCT Conference on Communication and Networking*, Kun Ming, China, Aug. 2012.

[HCAC15] H. Holm, A. Checko, R. Al-obaidi, and H. Christiansen, "Optimal Assignment of Cells in C-RAN Deployments with Multiple BBU Pools", in *Proc. of EuCNC'15 – 24th European Conference on Networks and Communications*, Paris, France, June 2015.

[HDDM13] M. Hadzialic, B. Dosenovic, M. Dzaferagic, and J. Musovic, "Cloud-RAN: Innovative Radio Access Network Architecture", in *Proc. of ELMAR'13 - 55th International Symposium of Electronics in Marine*, Zadar, Croatia, Sep. 2013.

[HGJL15] B. Han, V. Gopalakrishman, L. Ji, and S. Lee, "Network Function Virtualization: Challenges and Opportunities for Innovations", *IEEE Communications Magazine*, Vol.53, No.6, Feb. 2015, pp. 90-97.

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

[HSMA14] H. Hawilo, A. Shami, M. Mirahmadi, and R. Assal, "NFV: State of the Art, Challenges, and Implementation in Next Generation Mobile Networks (vEPC)", *IEEE Network Magazine*, Vol. 28, No. 6, Nov. 2014, pp. 18-26.

[INEs11] Instituto Nacional de Estatística, *Census 11 -: Resident population, present population, families, households, dwellings and buildings* (in Portuguese), Public Consultation, Portugal, 2011 [Online]. Available: http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos_quadros [Accessed: 26th Sept. 2016].

[JSSA14] M. Jammal, T. Singh, A. Shami, R. Asal, and Y. Li, "Software-Defined Networking: State of the Art and Research Challenges", *IEEE Communications Magazine*, vol.52, No.6, June 2014, pp. 210-217.

[JZHT14] M. Jarschel, T. Zinner, T. Hoßfeld, P. Tran-Gia, and W. Kellerer, "Interfaces, Attributes and Use Cases – A Compass for SDN", *IEEE Communications Magazine*, Vol.52, No.6, June 2014, pp. 210-217.

[LKKu99] W. Leinberger, G. Karypis, and V. Kumar, "Multi-capacity bin packing algorithms with

applications to job scheduling under multiple constraints", *in Proc. of ICPP 1999 – International Conference on Parallel Processing*, Aizu-Wakamatsu City, Japan, Set. 1999.

[LLXB12] L. Li, J. Liu, K. Xiong, and P. Butovitsh, "Field Test of Uplink CoMP Joint Processing with C-RAN Testbed", in *Proc. of CHINACOM 2012 – 7th International ICST Conference on Communication and Networking*, Kun Ming, China, Aug. 2012.

[LSJR13] C. Liu, K. Sundaresan, M. Jiang, S. Rangarajan, and G. Chang, "The Case for Re-Configurable Backhaul in Cloud-RAN based Small Cell Networks", in *Proc. INFOCOM 2013 - 32nd IEEE International Conference on Computer Communications*, Turin, Italy, Apr. 2013.

[LZGN14] J. Liu, S. Zhou, J. Gong, and Z. Niu," On the Statistical Multiplexing Gain of Virtual Base Station Pools", in *Proc. of GLOBECOM 2014 - IEEE Global Communications Conference,* Austin, TX, USA, Dec. 2014.

[Maro15] A. Marotta, *Optimisation of Radio Access Network Cloud Architectures Deployment in LTE-Advanced*, M.Sc. Thesis, IST, University of Lisbon, Lisbon, Portugal, 2015.

[MBCT15] F. Musumeci, C. Bellanzon, N. Carapellese, M. Tornatore, A. Pattavina, and S. Gosselin, "Optimal BBU Placement for 5G C-RAN Deployment Over WDM Aggregation Networks", *Journal of Lightwave Technology*, Vol. 34, No. 8, April 2015, pp. 1963-1970.

[MSGB15] R. Mijumbi, J. Serrat, J. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network Function Virtualisation: State-of-the-art and Research Challenges", *IEEE Communications Surveys and Tutorials*, Vol. 8, No. 1, Sep. 2015, pp. 236-262.

[NaWK12] S. Namba, T. Warabino, and S. Kaneko, "BBU-RRH Switching Schemes for Centralized RAN", in *Proc. of CHINACOM 2012 – 7th ICST Conference on Communication and Networking*, Kun Ming, China, Aug. 2012.

[NGMN13] NGMN, *Suggestions on Potential Solutions to C-RAN*, Report, January 2013 [Online]. Available: https://www.ngmn.org/uploads/media/NGMN_CRAN_Suggestions_on_Potential_Solutions_to_CRAN.pdf [Accessed: Oct. 10th 2015].

[NNBK13] T. Nakamura, S. Nagata, A. Benjebbour, Y. Kishiyama, T. Hai, S. Xiaodong, Y. Ning, and L. Nan, "Trends in Small Cell Enhancements in LTE Advanced", *IEEE Communications Magazine*, Vol.51, No.2, Feb. 2013, pp. 98-105.

[ONFo13] Open Networking Foundation, *Software-Defined Networking: The New Norm for Networks*, White Paper, Apr. 2012.

[Open10] Openet, *Closing the Mobile Data Revenue Gap*, White Paper, 2010.

[PCCR13] A. Pizzinat, P. Chanclou, F. Le Clech, T.L. Reedeker, Y. Lagadec, F. Saliou, and P. Galli, "Optical fibre solution for mobile fronthaul to achieve Cloud Radio Access Network" in *Proc. of FuNeMS 2013 – 22nd Future Network and Mobile Summit*, Lisbon, Portugal, July 2013.

[PWLP15] M. Peng, C. Wang, V. Lau, and V. Poor, "Fronthaul-Constrained Cloud Radio Access

Networks: Insights and Challenges", *IEEE Wireless Communications*, Vol.22, No.2, Apr. 2015, pp.152-160.

[SASR15] K. Sundaresan, M. Arslan, S. Singh, S. Rangarajan, and S. Krishnamurthy, "FluidNet: A Flexible Cloud-Based Radio Access Network for Small Cells", in *Proc. of ACM 19th International Conference on Mobile Computing & Networking*, New York, NY, USA, Sep. 2013.

[SeTB11] S. Sesia, I. Toufik, and I. Baker, *LTE. The UMTS Long Term Evolution: From Theory to Practice (2nd edition)*, John Wiley & Sons, Chichester, United Kingdom, 2011.

[Silv16] H. Silva, *Design of C-RAN Fronthaul for Existing LTE Networks*, M.Sc. Thesis, IST, University of Lisbon, Lisbon, Portugal, 2016.

[SSCF13] S. Sezer, S. Scott-Hayward, P.K. Choulan, B. Fraser, D. Lake, J. Finnegan, N. Viljoen, M. Miller, and N. Rao "Are we ready for SDN? – Implementation Challenges for Software-Defined Networks", *IEEE Communications Magazine*, Vol.51, No.7, July 2013, pp. 36-43.

[SuRF15] V. Suryaprakash, P. Rost, G- Fettweis, "Are Heterogenous Cloud-Based Radio Access Networks Cost Effective?", *IEEE Journal on Selected Areas in Communications*, Vol. 33, No. 10, Oct. 2015, pp. 2239-2251.

[WeGP13] T. Werthman, H. Grob-Lipski, and M. Proebster, "Mutiplexing Gains Achieved in Pools of Baseband Computation Units in 4G Cellular Networks", in *Proc. of PIMRC 2013 – 24th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications*, London, United Kingdom, Sep. 2013.