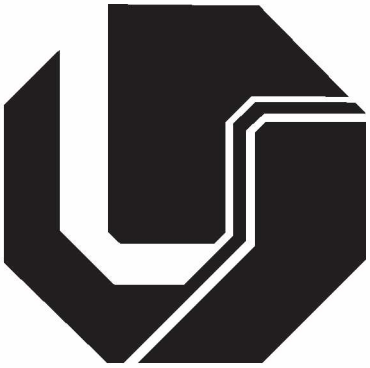
FEDERAL UNIVERSITY OF UBERLÂNDIA
FACULTY OF ELECTRICAL ENGINEERING



A NEW GENETIC ALGORITHM BASED SCHEDULING ALGORITHM FOR THE LTE UPLINK

SAULO HENRIQUE DA MATA

Uberlândia - 2017

SAULO HENRIQUE DA MATA

A NEW GENETIC ALGORITHM BASED SCHEDULING ALGORITHM FOR THE LTE UPLINK

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Sciences to the Post-Graduate Program of the Faculty of Electrical Engineering at the Federal University of Uberlândia.

Members of the Committee:

Prof. Dr. Paulo Roberto Guardieiro (Advisor) - UFU
Profa. Dra. Juliana Freitag Borin - UNICAMP
Prof. Dr. Márcio Andrey Teixeira - IFSP
Prof. Dr. Éderson Rosa da Silva - UFU
Prof. Dr. Márcio José da Cunha - UFU

Uberlândia - 2017

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

M425n
2017

Mata, Saulo Henrique da, 1986-
A new genetic algorithm based scheduling algorithm for the LTE Uplink / Saulo Henrique da Mata. - 2017.
120 f. : il.

Orientador: Paulo Roberto Guardieiro.
Tese (doutorado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Engenharia Elétrica.
Inclui bibliografia.

1. Engenharia elétrica - Teses. 2. Long-Term Evolution (Telecomunicações) - Teses. 3. Algoritmos genéticos - Teses. I. Guardieiro, Paulo Roberto, 1952- II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU: 621.3

SAULO HENRIQUE DA MATA

A NEW GENETIC ALGORITHM BASED SCHEDULING ALGORITHM FOR THE LTE UPLINK

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Sciences to the Post-Graduate Program of the Faculty of Electrical Engineering at the Federal University of Uberlândia.

Prof. Paulo Roberto Guardieiro, Dr.	Prof. Alexandre Cardoso, Dr.
Advisor	Coordinator of the Post-Graduation Program

Acknowledgments

Firstly, I would like to thank God, that was with me in each day of work. Thank You for the perseverance, patience and health, that were present during this journey.

To prof. Dr. Paulo Roberto Guardieiro for the always dedicated and patient guidance. Thank you for the encouragement and for teaching me not only technical concepts, but also for being an example of conduct and professionalism.

To my family, for the support at all times. I thank my mother Domingas, my father José and my fiancée Joyce for their support in difficult times, for understanding my absence in so many moments and for the trust placed in me.

Thanks to my friends for the company, support, learning, and so many moments of relaxation and joy.

Finally, I thank FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais) for the financial assistance to make this work viable.

"I can do all things through Christ which strengtheneth me."
Philippians 4,13

Abstract

da Mata, S. H., *A NEW GENETIC ALGORITHM BASED SCHEDULING ALGORITHM FOR THE LTE UPLINK* , UFU, Uberlândia, Brazil, 2017, 102p.

Long Term Evolution has become the *de facto* technology for the 4G networks. It aims to deliver unprecedented data transmission rates and low latency for several types of applications and services. In this context, this thesis investigates the resource allocation in the LTE uplink. From the principle that resource allocation in the uplink is a complex optimization problem, the main contribution of this thesis is a novel scheduling algorithm based on Genetic Algorithms (GA). This algorithm introduces new operations of initialization, crossover, mutation and a QoS-aware fitness function. The algorithm is evaluated in a mixed traffic environment and its performance is compared with relevant algorithms from the literature. Simulations were carried out in ns-3 and the results show that the proposed algorithm is able to meet the Quality of Service (QoS) requirements of the applications, while presenting a satisfactory execution time.

Index-terms: LTE, Uplink, Scheduling Algorithms, Genetic Algorithms, ns-3

Contents

List of Figures

List of Tables

List of Abbreviations

16-QAM	16-Quadrature Amplitude Modulation
1G	First-Generation Mobile Networks
2G	Second-Generation Mobile Networks
3G	Third-Generation Mobile Networks
3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
4G	Fourth-Generation Mobile Networks
64-QAM	64-Quadrature Amplitude Modulation
AMBR	Aggregate Maximum Bit Rate
AMC	Adaptive Modulation and Coding
AMPS	Analogue Mobile Phone System
ARP	Allocation and Retention Priority
AS	Access Stratum
ATB	Adaptive Transmission Bandwidth
BCCH	Broadcast Control Channel
BCH	Broadcast Channel
BLER	Block Error Rate
BSR	Buffer Status Report
CA	Carrier Aggregation
CAC	Call Admission Control

CCCH	Common Control Channel
CCI	Co-Channel Interference
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CFI	Control Format Indicator
CoMP	Coordinated Multi-Point
CQI	Channel Quality Indicator
CSI	Channel State Information
CTTC	Centre Tecnològic de Telecomunicacions de Catalunya
DCCH	Dedicated Control Channel
DCI	Downlink Control Information
DFTS-OFDMA	Discret Fourier Transform Spread Orthogonal Frequency Division Multiple Access
DL-SCH	Downlink Shared Channel
DTCH	Dedicated Traffic Channel
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
EDGE	Enhanced Data Rates for GSM Evolution
EMM	EPS Mobility Management
eNB	evolved NodeB
EPA	Extended Pedestrian A
EPC	Evolved Packet Core
EPS	Evolved Packet System
ESM	EPS Session Management
ETSI	European Telecommunications Standards Institute
EV-DO	Evolution Data Optimized
EVA	Extended Vehicular A

FDD	Frequency Division Duplexing
FDPS	Frequency Domain Packet Scheduling
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FME	First Maximum Expansion
FTP	File Transfer Protocol
GA	Genetic Algorithms
GBR	Guaranteed Bit Rate
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
GTP-U	GPRS Tunneling Protocol User Part
GUI	Graphical User Interface
HARQ	Hybrid Automatic Repeat Request
HI	Hybrid ARQ Indicator
HRPD	High Rate Packet Data
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
HSUPA	High Speed Uplink Packet Access
IMT-2000	International Mobile Telecommunications-2000
IMT-2020	International Mobile Telecommunications-2020
IMT-Advanced	International Mobile Telecommunications-Advanced
IP	Internet Protocol
ISI	Inter Symbol Interference
ITU	International Telecommunication Union
ITU-R	International Telecommunication Union Radiocommunication Sector

J-TACS	Japanese Total Access Communication System
LA	Link Adaptation
LENA	LTE-EPC Network simulAtor
LTE	Long-Term Evolution
MAC	Medium Access Control
MAD^E	Minimum Area-Difference to the Envelope
MBMS	Multimedia Broadcast/Multicast Service
MBR	Maximum Bit Rate
MCH	Multicast Channel
MCS	Modulation Coding Scheme
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
MT	Maximum Throughput
MT	Mobile Termination
MTC	Machine-Type Communication
NAS	Non Access Stratum
NMT	Nordic Mobile Telephone System
Non-GBR	Non-Guaranteed Bit-Rate
ns-3	Network Simulator 3
OFDMA	Orthogonal Frequency Division Multiple Access
P-GW	PDN Gateway
PAPR	Peak-to-Average Power Ratio
PBCH	Physical Broadcast Channel
PCCH	Paging Control Channel
PCFICH	Physical Control Format Indicator Channel
PCH	Paging Channel

PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PDSCH	Physical Downlink Shared Channel
PDU	Protocol Data Unit
PF	Proportional Fairness
PHICH	Physical Hybrid ARQ Indicator Channel
PHR	Power Headroom Report
PLR	Packet Loss Ratio
PMCH	Physical Multicast Channel
PMI	Pre-coding Matrix Indicators
PRACH	Physical Random Access Channel
PS	Packet Scheduler
PSNR	Peak Signal-to-Noise Ratio
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QCI	QoS Class Identifier
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RA	Resource Allocation
RACH	Random Access Channel
RB	Resource Block
RBG	Radio Bearer Group
RI	Rank Indications
RLC	Radio Link Control
RME	Recursive Maximum Expansion

RP	Riding Peaks
RR	Round Robin
RRC	Radio Resource Protocol
RRM	Radio Resource Management
S-GW	Serving Gateway
S1-AP	S1 Application Protocol
SAE	System Architecture Evolution
SAW	Stop-And-Wait
SC-FDMA	Single Carrier Frequency Division Multiple Access
SCTP	Stream Control Transmission Protocol
SIC	Successive Interference Cancellation
SINR	Signal to Interference plus Noise Ratio
SMS	Short Message Service
SNR	Signal to Noise Ratio
SR	Scheduling Requests
SRS	Sounding Reference Signals
TACS	Total Access Communications System
TB	Transport Block
TCP	Transmission Control Protocol
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TDPS	Time Domain Packet Scheduling
TE	Terminal Equipment
TEID	Tunnel Endpoint Identifier
TFT	Traffic Flow Template
TTI	Transmission Time Interval

UCI	Uplink Control Information
UDP	User Datagram Protocol
UE	User Equipment
UICC	Universal Integrated Circuit Card
UL-SCH	Uplink Shared Channel
UMTS	Universal Mobile Telecommunication System
USIM	Universal Subscriber Identity Module
UTRAN	UMTS Terrestrial Radio Access Network
VoIP	Voice over IP
WCDMA	Wideband Code Division Multiple Access

"Science is organized knowledge. Wisdom is organized life."

- Immanuel Kant

1

Introduction

In the past years, we have been witnessing a tremendous growth of mobile networks subscribers. This fact has demanded a continuous evolution of the current mobile networks, to attend the always crescent expectations of these users. Applications, such as Voice over IP (VoIP), web browsing, video chat and video streaming, have applied new challenges to the design of mobile networks, because of their delay and bandwidth strict requirements.

In this context, the Long-Term Evolution (LTE) network has emerged as one of the most promising solutions to overcome these new challenges. LTE is a packet-based mobile broadband network. It has been developed by the Third Generation Partnership Project (3GPP) and aims to deliver high throughput, low latency and an enhanced spectral efficiency with respect to previous 3G networks [1].

As stated before, one can find a plurality of different applications, each of them with specific delay and bandwidth requirements. These delay and bandwidth requisites can be mapped into different Quality of Service (QoS) classes. Therefore, it is of fundamental importance that the network can ensure these QoS requirements for each of those applications. Generally, the Call Admission Control (CAC) and the Resource Allocation (RA) are the main mechanisms to ensure these QoS requisites.

In this sense, the resource allocation mechanism is a key feature of the LTE network. However, the plurality of applications and different QoS requirements bring to light complex challenges to resource allocation design.

In the next sections, we identify these main challenges, and what have been developed by the community of researchers. This will allow us to build a map to guide our efforts in the proposal of a new scheduling algorithm for the LTE network.

1.1 Problem Definition

In the LTE system, the User Equipment (UE) gets access to the network through the base station, which is known as evolved NodeB (eNodeB or eNB). The eNodeB is responsible for the allocation of the network resources among the UEs attached to it.

In a real propagation environment, the air interface is characterized by fast fading variations, resulting from the multiple possible paths that the signal can travel until it reaches the receiver. Depending on the path, it can occur a constructive or a destructive recombination of the signal at the receiver. The position and distance of the receiver from the transmitter also influence these fast fading variations. Moreover, high data rate transmission in a multipath environment leads to Inter Symbol Interference (ISI) and, consequently, bit errors at the receiver [2].

From the previous 3GPP mobile networks, one of the most important changes introduced by LTE is the shift from the use of Code Division Multiple Access (CDMA) to Orthogonal Frequency Division Multiple Access (OFDMA) [3]. OFDMA simplifies the design of channel equalizers and it is a powerful way to solve the ISI problem. Furthermore, OFDMA offers high spectral efficiency, scalability and flexibility of bandwidth allocation, since the resource allocation can occur in time and frequency domains.

Despite the great advantages of OFDMA, it also presents some issues. The main one is the high Peak-to-Average Power Ratio (PAPR). This means that the power of the transmitted signal is subject to rather large variations, which cause the amplifiers to reach saturation region, resulting in a non-linear distortion. In the downlink, the eNodeB transmitters are large, expensive devices, so they can avoid this problem by using expensive power amplifiers that are very close to linear. In the uplink, a UE transmitter has to be cheap and present reduced power consumption [2].

In this context, LTE uses OFDMA in the downlink, but for the uplink, LTE uses a variant of the OFDMA, which is known as Single Carrier Frequency Division Multiple Access (SC-FDMA). SC-FDMA presents most of the benefits of OFDMA and it also

offers reduced power consumption and improved coverage. On the other hand, it requires the subcarriers allocated to a single UE to be adjacent [4]. Therefore, the scheduler for the uplink has limited degrees of freedom: it has to allocate contiguous Resource Blocks (RBs) to each user without the possibility of choice among the best available ones [1]. This constraint will prove to be very challenging when designing scheduling algorithms [4], since the incorporation of the RB contiguity constraint into the uplink scheduling algorithms was proven to be NP-hard [5], i.e., it is impractical to perform an exhaustive search.

Downlink and uplink have a radio link with a time variant nature, due to the fast fading phenomenon, as said before. Thus, the eNodeB must consider the UEs current quality of channel to allocate the resources in an effective manner. In the downlink, the quality of the channel is obtained through the Channel Quality Indicator (CQI), which is reported by the UE to the eNodeB. For the uplink, the eNodeB estimates the channel's quality using channel sounding techniques. From the channel's quality, the system can perform link adaptation using Adaptive Modulation and Coding (AMC) techniques, i.e. the system can choose a more robust Modulation Coding Scheme (MCS) under adverse channel conditions to improve spectral efficiency.

In this sense, channel-aware solutions are usually adopted in LTE resource allocation, since they are able to exploit channel quality variations by assigning higher priority to users experiencing better channel conditions. However, the channel quality cannot be the only factor in the resource scheduling process. The scheduling algorithm must also take into consideration, for example, the average throughput of the cell, fairness index and mainly the QoS requirements.

Finally, one can see that the design of a scheduling algorithm for the LTE network is a complex task. There are many issues to be addressed in order to obtain a high spectral efficiency and to meet the QoS requirements. Thus, a powerful and effective scheduling algorithm should be channel-aware/QoS-aware. The uplink channel is even more challenging, since the SC-FDMA requires that the subcarriers allocated to a particular UE must be adjacent, elevating the complexity to NP-hard level [5].

In the next section, the most relevant works that propose solutions for the above issues are presented.

1.2 State of the Art

Resource allocation in the LTE networks has been a field of intense research. The downlink channel has concentrated the majority of studies. Capozzi *et al.*[1] enumerate

the key features for the downlink packet scheduler design and present a survey with the main papers about this issue.

The uplink channel has received less attention, but one can find interesting papers dealing with the uplink scheduling challenges, as listed in [6]. In [7], we can find one of the first proposals to solve the problem of the localization constraint for RB assignment. It is based on an earlier work [8], which suggests assigning RBs to users who obtain the highest marginal utility.

Calabrese *et al.* [9] propose a search-tree algorithm, assuming the resources equally shared among users. Further, Calabrese also proposed in [10] an Adaptive Transmission Bandwidth (ATB) algorithm. According to the authors, the main advantage provided by ATB scheduling algorithm is more flexibility to accommodate the varying cell load, when compared to the search-tree algorithm.

Temiño *et al.* [11] propose three new scheduling algorithms: the First Maximum Expansion (FME), the Recursive Maximum Expansion (RME) and the Minimum Area-Difference to the Envelope (MAD^E). According to authors, RME outperforms FME. MAD^E presents a small gain over RME in some scenarios, but with a much greater computational complexity.

In another contribution [5], Lee *et al.* propose four scheduling algorithms. Among them, we highlight the Riding Peaks (RP) algorithm. It is based on the fact that in multi-carrier systems, the channel state of a user is correlated in both time and frequency. Thus, it tries to use each user's highest valued RBs as much as possible. This algorithm is simple and performs quite well, but according to the authors it leads to bad solutions if the channel presents abrupt changes of the Signal to Noise Ratio (SNR) from one RB to the next.

Safa *et al.* [12] performe a comparison of three scheduling algorithms: RME, FME and RP. Despite of the issues indicated in [5], according to Safa, the RP algorithm presented better performance than the other two algorithms under evaluation.

In [13], Kaddour *et al.* also present a comparison of RP, RME and their own proposed algorithm in the context of a power efficient resource allocation. As expected, RP and RME did not perform well in these scenarios, since they do not consider the transmission power in the resource allocation.

So far, the aforementioned scheduling algorithms were evaluated with only one type of traffic source and they do not have any kind of mechanisms to assure the QoS requirements. In [14], one can find another comparative evaluation with a set of scheduling algorithms. However, this time, the comparison is performed with mixed traffic, i.e.,

considering VoIP, video streaming and FTP applications. The authors highlight the challenges to provide a fair evaluation environment for comparing the schedulers proposed for the uplink. They also affirm that the schedulers under evaluation performed almost similarly.

From the algorithms presented in [5], [9] and [11], we can find a set of works that tried to improve these three solutions and add QoS capabilities, by giving different priorities to the traffics from the utility function, as in [15], [16] and [17]. In [15], Safa *et al.* present an interesting scheduler that will be described in Chapter 4 and used in the performance evaluation in Chapter 6.

As stated in Section 1.1, finding the optimal solution to the LTE uplink scheduling is a NP-hard problem. In this sense, greedy heuristic algorithms can find "good enough" solutions, performing a trade-off between computational performance and accuracy. At their worst case, they might not perform well, but in practice, their overall performance is very good [12]. On the other hand, there are tools that were designed to deal with NP-hard problems. One of the most known is the Genetic Algorithms (GA). GA is capable of delivering near-optimal performance with comparatively low complexity [18].

GA has been used in scheduling algorithms mainly for the downlink, as in [3] - [20]. For the uplink, one can find a set of works dealing with mitigation of Co-Channel Interference (CCI) using Successive Interference Cancellation (SIC) as described in [21] and [22]. These works are more related to Coordinated Multi-Point (CoMP) transmission and multi user scheduling. It does not consider the NP-hard issue of the localization constraint for the RB assignment.

In [23], Kalil *et al.* also employ GA in the LTE uplink scheduling and they considered the RB contiguity constraint. However, the proposed algorithm is compared only with the optimal solution, and the authors used a simple scenario with a small number of users and there is not any information about the traffic in the cell.

1.3 Justification

Based on what was presented in the previous sections, one can see that the resource allocation in LTE uplink is a complex task, rated as NP-hard. Despite the availability of tools to solve this kind of problem, like the Genetic Algorithms, to the best of our knowledge, there are few studies about the use of GA as the heuristic of an LTE uplink scheduling algorithm, regarding the contiguity RB constraint.

Considering the power and history of GA to solve this kind of problem, and taking

also into account the several already existent uplink scheduling algorithms based on traditional techniques, we believe that there is a great opportunity of research in the design of an LTE uplink scheduling based on the Genetic Algorithms.

1.4 Research Scope and Objectives

The main objective of this work is to solve the resource allocation problem in the LTE uplink. In order to achieve this goal, we considered the complexity of this optimization problem and chose the Genetic Algorithms as the strategy of scheduling.

GA is powerful tool to solve complex optimization problems and has not been fully studied in this context yet. In this sense, a second objective of this work is to verify if GA can be applied in the LTE uplink resource allocation and how it performs compared to the literature solutions. Due to restrictions of the simulation tool, this thesis will focus on the Release 8 of the LTE standard. When suitable, we point out the impact of features of newer releases in the proposal described in this thesis. The optimization of the GA parameters are not in the scope of this thesis. On the other hand, the proposed algorithm should be able to meet the QoS requirements of different applications, considering the contiguity constraint for RB assignment, while presenting a satisfactory execution time.

To evaluate the proposed scheduling algorithm it is necessary to define a simulation environment, since there are a substantial number of parameters involved, resulting into an very complex analytical approach, and a real testbed would be very expensive. From the papers mentioned in Section 1.2, one can see that most of the authors use their own implementation of simulator or commercial/private simulators. In summary, these simulation environments are neither publicly available nor complete nor easy to use. This brings a huge challenge for researchers interested in studying and comparing their works with the state of art solutions. This fact was highlighted by the authors of [14].

In this context, a third objective of this work is to make available the results and tools developed to evaluate our scheduling algorithm. It is worth saying that we will not develop a new simulation environment. Instead, we will use an already existent and popular simulator and develop the modules and tools necessary to evaluate our proposal. This is a very important step to facilitate and promote future works based on the scheduling algorithm presented in this thesis.

1.5 Novelty and Contributions

The main contribution of this thesis is a novel scheduling algorithm based on Genetic Algorithms for the LTE uplink networks.

From the scheduling algorithms listed in Section 1.2, the only one that is really close to our proposal is the one described in [23]. The closeness can be noted in the strategy to solve the resource allocation problem, i.e., both schedulers use Genetic Algorithms to allocate resources in the uplink channel, considering the contiguity constraint. However, representation, initialization, crossover, mutation and evaluation are distinct in the two proposals. When compared to [23], our proposal presents the following improvements:

- the algorithms are evaluated using three traffic models for FTP, VoIP and Video Chat applications, while the evaluation in [23] presents only one traffic;
- our proposal presents a new three-step strategy of allocation. We believe that this three-step approach is the main element responsible for making our algorithm competitive in execution time. In [23], there is not a complexity analysis;
- our algorithm presents a novel QoS-aware fitness function. The fitness function in [23] is not QoS-aware;
- our scheduler employs a packet discarding strategy to meet the packet delay budget of the applications;
- our evaluation presents a fairer, challenger and more realistic comparison, since we added a QoS-aware algorithm in the evaluation to compete with our proposal.

Furthermore, the scheduling algorithm presented in this thesis, introduces new strategies of initialization, crossover and mutation that complies with the contiguity constraint for the RBs assignment.

To the best of our knowledge, there is not any previous work that proposes such a robust GA-based scheduler and presents a complete evaluation, considering several network performance indicators.

It is also worth saying that to evaluate the proposed algorithm, we had to implement new scheduling algorithms and traffic generators in ns-3 [24]. These implementations can help other researchers in their evaluations.

To facilitate the conduction of the simulations and the gathering of results, we developed a Graphical User Interface (GUI) for the LTE module of ns-3 [24]. This tool is in beta state, but we are very sure that it can be useful for researches in this field.

Finally, we would like to highlight that the novelty and contributions of the research presented in this thesis were corroborated and accepted by the community of researchers, by means of the following publications:

- S. H. da Mata, P. R. Guardieiro, Resource allocation for the LTE uplink based on Genetic Algorithms in mixed traffic environments, Computer Communications 107 (2017) 125–137. doi:10.1016/j.comcom.2017.04.004.
- J. M. H. Magalhães, S. H. da Mata, P. R. Guardieiro, Downlink and Uplink Resource Allocation in LTE Networks, in: Handbook of Research on Next Generation Mobile Communication Systems, IGI Global, 2015, pp. 199–233. doi: 10.4018/978-1-4666-8732-5.ch009.
- S. H. da Mata, P. R. Guardieiro, A Genetic Algorithm Based Approach for Resource Allocation in LTE Uplink, in: 2014 International Telecommunications Symposium (ITS), 2014, pp. 1–5. doi:10.1109/ITS.2014.6947956.

1.6 The Structure of the Thesis

This thesis is organized as follows.

Chapter 2 is dedicated to the introduction of the main concepts of the LTE network that are important to understand the research described in this thesis.

In Chapter 3, we give more details of the resource allocation process in the LTE system. This chapter describes how the LTE resources are organized to be shared among the users.

Chapter 4 is dedicated to present the packet scheduler. In this chapter, we introduce the main requirements and functionalities of the packet scheduler and present the most relevant scheduling algorithms found in the literature.

In Chapter 5, we describe the proposed algorithm. This chapter introduces the Genetic Algorithms and how this technique is used to solve the LTE uplink scheduling issues.

In Chapter 6, we evaluate the proposed algorithm. This chapter also describes the simulation environment used to perform the evaluation and it presents the results of the evaluation as well.

Finally, Chapter 7 presents the conclusions and future developments of the research.

"Never memorize something that you can look up."

- Albert Einstein

2

The Long-Term Evolution Network

The Long-Term Evolution is a very complex network with many elements that make use of several concepts and technologies. This chapter presents some of these elements and concepts that are important to facilitate the understanding of the research delineated in this thesis.

2.1 Historical Context of the Mobile Networks

LTE is commonly known as a mobile network of fourth generation (4G) [2]. In this sense, there is a logical thought about what previous mobile networks technologies formed the past generations. In this section, we have a brief discussion about the evolution of mobile networks.

2.1.1 First-Generation Mobile Networks (1G)

The First-Generation of mobile networks arrived in the 1980s. There was a set of independent projects around the globe [25]:

- The Analogue Mobile Phone System (AMPS), used in the USA.
- The Nordic Mobile Telephone System (NMT) and the Total Access Communications System (TACS), used in Europe.
- The Japanese Total Access Communication System (J-TACS), used in Japan and Hong Kong.

All these systems were based on analogue communication technologies and limited only to voice services [26]. Despite the good voice quality, these systems presented a limited spectral efficiency. Besides, the mobile devices were large and expensive. These problems motivated the development of a new mobile network.

2.1.2 Second-Generation Mobile Networks (2G)

In the early 1990s, the digital communications were already mature enough to support commercial systems. In this context, a Second-Generation of mobile networks were developed based on digital technology, which permitted a better spectral efficiency, associated with smaller and cheaper devices [2].

Initially, these systems were developed to support only voice services. However, later, they were enhanced to support circuit-switched data services. One of these first services was the Short Message Service (SMS) [27].

The most popular 2G system is the Global System for Mobile Communication (GSM), designed by a set of companies working together under the guidance of the European Telecommunications Standards Institute (ETSI). GSM uses Time Division Multiple Access (TDMA) as the transmission method.

Another notable standard is IS-95, also known as *cdmaOne*, designed by Qualcomm and used in the USA. This system is based on Code Division Multiple Access (CDMA).

The small and long battery life terminals of 2G networks helped to spread these technologies worldwide. Simultaneously, Internet was presenting a very promising growth. This context promoted the junction of these two concepts: Internet data services offered in a mobile device. To achieve this goal, network operators developed enhancements in the original 2G networks. These enhanced systems are often referred as 2.5G mobile networks [27]. The main improvements were the introduction of a packet-switched core network and a modified air interface to handle voice and data transmission.

The GSM system incorporated these improvements with the development of the General Packet Radio Service (GPRS). GPRS reached peak data rates up to 140 Kbps

[26]. Further enhancements were integrated to the GSM network with the introduction of the Enhanced Data Rates for GSM Evolution (EDGE) technique. EDGE has increased data rates up to 384 Kbps [26].

2.1.3 Third-Generation Mobile Networks (3G)

In 1997, the International Telecommunication Union (ITU) published the Recommendation ITU-R M.687-2 entitled International Mobile Telecommunications-2000 (IMT-2000) [28]. This recommendation was launched as an initiative to promote high quality mobile multimedia networks, which can offer high speed broadband, Internet Protocol (IP) based system and global roaming, among other features.

From this initiative, two partnership organizations were born: the Third Generation Partnership Project (3GPP) and the Third Generation Partnership Project 2 (3GPP2). Each of these organizations developed their own technologies for IMT-2000 specifications.

The most popular 3G system in the world is the Universal Mobile Telecommunication System (UMTS) [2]. UMTS is a system developed from the GSM by the 3GPP. They changed the air interface, while kept the core network almost the same. The radio access, known as UMTS Terrestrial Radio Access Network (UTRAN), is based on Wideband Code Division Multiple Access (WCDMA).

On the other hand, 3GPP2 developed *cdma2000* from *cdmaOne*. There are three main differences between UTRAN and *cdma2000* systems [2]:

- UTRAN uses a bandwidth of 5 MHz, while *cdma2000* uses 1.25 MHz.
- *cdma2000* is backwards compatible with *cdmaOne*, i.e., *cdmaOne* devices can communicate with *cdma2000* base stations and conversely. This is not true for UMTS and GSM.
- UTRAN allows the carrier frequencies to be shared by voice and data traffic. On the other hand, *cdma2000* segregates voice and optimized data onto different carrier frequencies.

Unfortunately, the performance of initial 3G systems did not meet the expectations. Hence, 3G only started to be promising after the introduction of 3.5G systems, in 2005.

3GPP brought out the High Speed Downlink Packet Access (HSDPA) and the High Speed Uplink Packet Access (HSUPA), which presented better spectral efficiency. Together, HSDPA and HSUPA are known as High Speed Packet Access (HSPA).

Finally, 3GPP presented the HSPA+ as the last evolution of HSPA, adding Multiple Input/Multiple Output (MIMO) antenna capability and 16-QAM (uplink)/64-QAM (downlink) modulations [26].

3GPP2 followed a similar direction and introduced the High Rate Packet Data (HRPD), also known as Evolution Data Optimized (EV-DO). Later, 3GPP2 enhanced their system and released CDMA EV-DO Revision A and CDMA EV-DO Revision B [26].

2.1.4 Fourth-Generation Mobile Networks (4G)

For many years, voice calls were the main traffic in mobile networks. According to [29], from 2011, data traffic started to grow dramatically. This scenario is depicted in Figure 2.1, which shows the total global data and voice traffic[1].

From Figure 2.1, we can see that voice traffic has a small growth tendency. On the other hand, data traffic has an exponential growth tendency. Part of this growth is owed to the success of 3.5G technologies and the spreading of smartphones and tablets with very attractive and easy to use applications.

This huge growth of data traffic started to congest the 2G and 3G networks. In this sense, it was necessary to find a strategy to increase the system capacity.

According to [2], from Shannon Limit equation, there are three ways to increase the system capacity:

- **The use of small cells**: The capacity of the base station is shared among the users. Using smaller cells, one has a greater number of base stations and a smaller number of users for each base station. It is worth noting that including more base stations means more costs to network operators.
- **Increasing the bandwidth**: with more bandwidth available, base station can handle a greater number of users. However, the radio spectrum is finite and it is under different regulations in each country.
- **Enhancement of communication technology**: the development of new technologies may allow the use of a higher Signal to Interference plus Noise Ratio (SINR) and better spectral efficiency.

The enhancement of communication technology is one of the main motivations for the development of the LTE network. We can also cite other three reasons [2]:

[1]Traffic does not include DVB-H, Wi-Fi, or Mobile WiMax. Voice does not include VoIP. M2M traffic is not included.

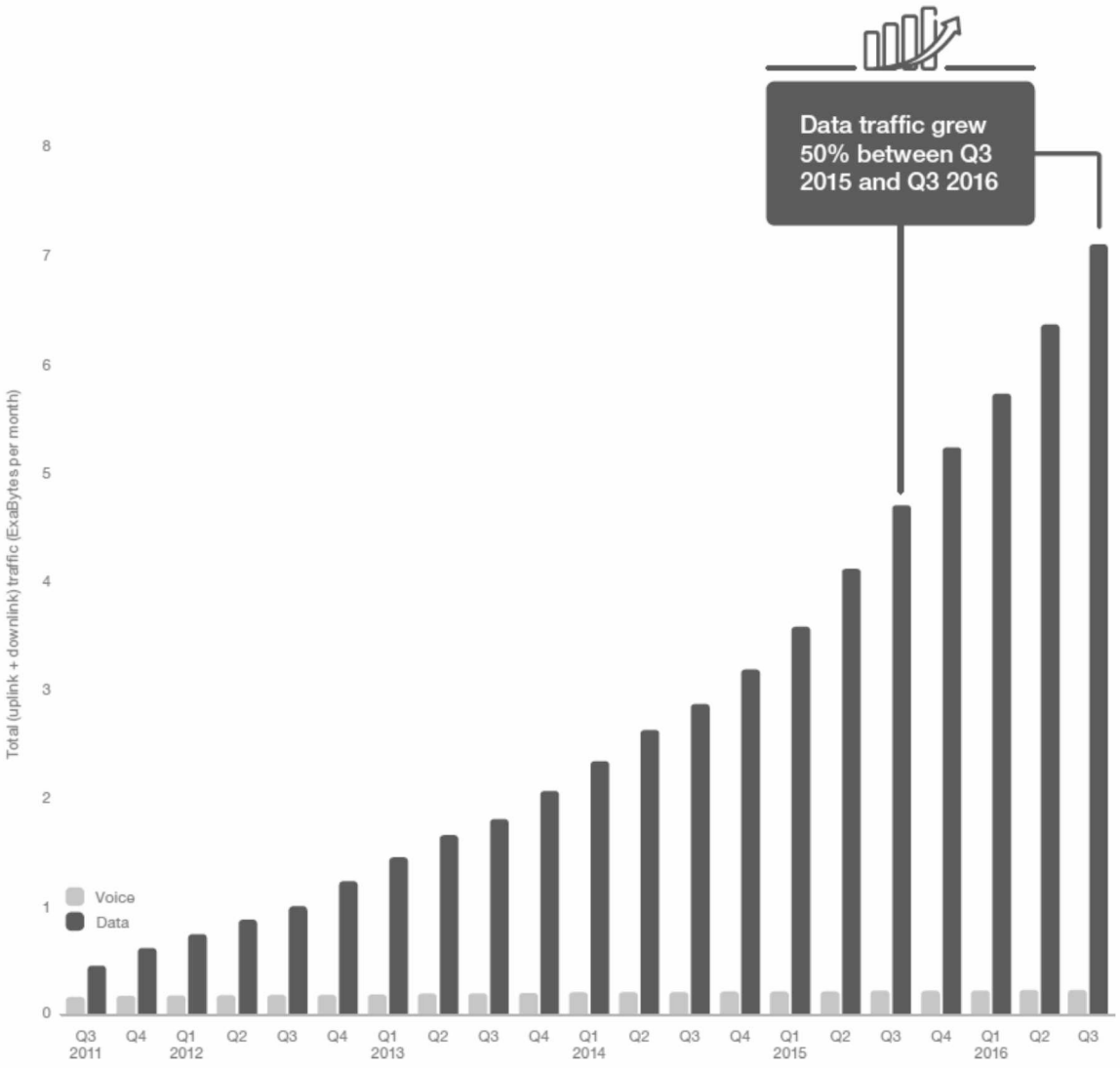


Figure 2.1: Global total data traffic in mobile networks, 2011-2016 (Adapted from [29]).

- 2G and 3G networks require two core networks: a circuit switched domain for voice and a packet switched domain for data. LTE has an all-IP design, i.e., only packet switched domain is used for voice and data.
- 3G networks introduce delays of the order of 100 milliseconds. This delay can be very prohibitive for real time applications. LTE was designed to provide a latency close to 10 milliseconds.
- It has been a complex task to add new features to UMTS and keep it backward compatible with GSM devices. LTE may represent a fresh start and facilitate the inclusion of new features to the network.

In this context, LTE has been developed to meet the demand for an improved system

with more capacity and to facilitate the management of the network by the operators.

As said before, LTE technology was designed as a fresh start, i.e., a packet-switching network developed to be all-IP, with no support to circuit-switched voice. The focus of the LTE is high data rates, low latency and high capacity [27].

An important aspect of the LTE network is its worldwide acceptance. Differently of previous generations, which there were competing technologies, LTE is a single technology, as depicted in Figure 2.2. This fact promoted the LTE spreading and accelerates the development of new services.

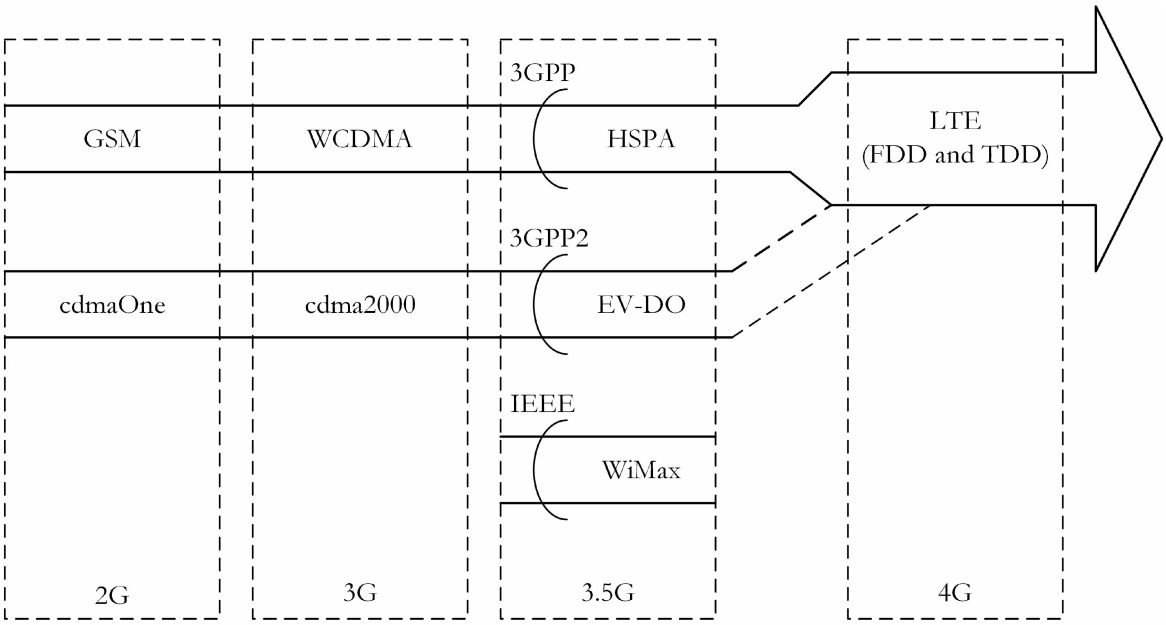


Figure 2.2: Convergence of wireless technologies (Adapted from [27]).

The first version of LTE was the Release 8. Since then, the standard has evolved through the years and new releases were developed and launched by the 3GPP. Release 10 is also known as LTE-Advanced, since it is the first release to fully meet the requirements of the IMT-Advanced specifications. The last launched release is Release 13, also known as LTE-Advanced Pro. This release is considered a 4.5G technology, i.e., a transition between the 4G and 5G technologies. As one can see, Figure 2.3 lists a set of features introduced by each LTE release. Next, the features that are correlated with our proposal will be briefly described.

OFDMA

LTE makes use of OFDMA to get high spectral efficiency. In downlink, OFDMA offers robustness against multipath interference and gives support to advanced techniques such

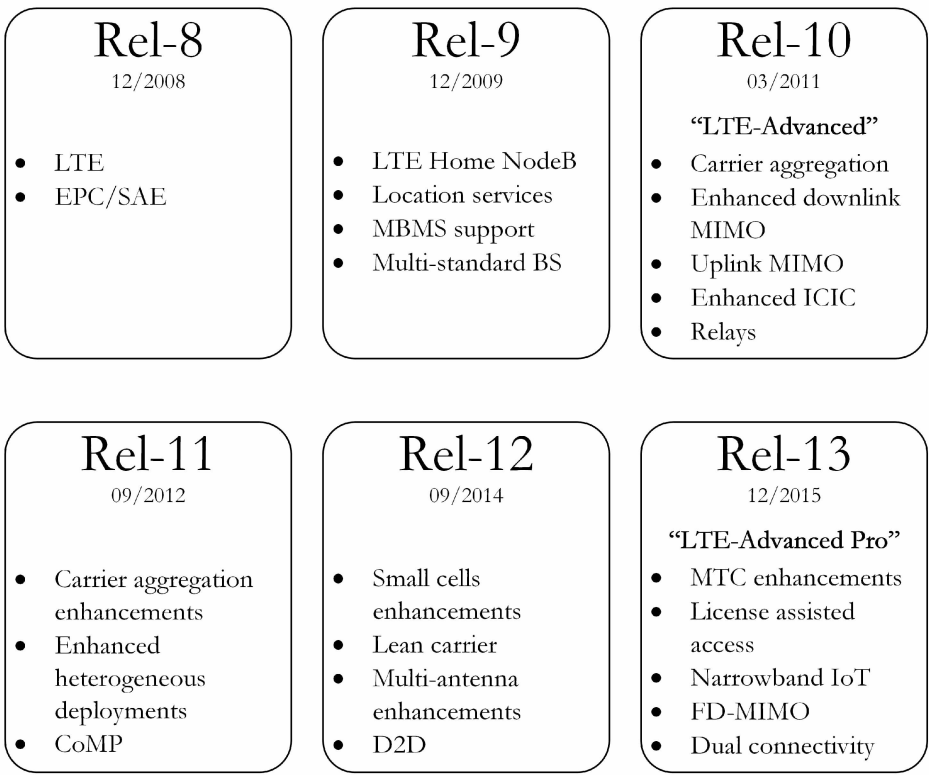


Figure 2.3: LTE releases (Adapted from [27]).

as frequency domain channel-dependent scheduling and MIMO techniques. However, uplink uses SC-FDMA, since it works with a lower PAPR. SC-FDMA also provides user orthogonality in frequency domain, and multi-antenna application [26].

Support for TDD and FDD

LTE supports Time Division Duplexing (TDD) and Frequency Division Duplexing (FDD). When using FDD, the base station and mobile transmit and receive simultaneously, but using different carrier frequencies. Using TDD, they transmit and receive on the same carrier frequency but at different time slots [2].

Adaptive Modulation and Coding (AMC)

AMC is an effective mechanism to maximize throughput in a time-varying channel. Using this technique, LTE can change the modulation and Forward Error Correction (FEC) coding schemes in a per user and per frame basis, depending on the channel conditions [26].

Support for variable bandwidth

Considering the Release 8/9, LTE supports operation in different bandwidth sizes: 1.4, 3, 5, 10, 15, and 20 MHz in both, uplink and downlink . This wide range of bandwidths gives flexibility for the network operators to deploy LTE in a variety of spectrum management regimes [2]. From Release 10, LTE started to support Carrier Aggregation (CA). In this strategy, the LTE network can aggregate multiple *component carriers* to increase the bandwidth. In Releases 10, 11 and 12, up to five component carriers can be aggregated, allowing the bandwidth to be increased up to 100 MHz. In Release 13, the specification defined up to 32 component carriers, elevating the aggregated bandwidth up to 640 MHz [27].

Very high peak data rates

LTE is capable of supporting very high peak data rates. In Releases 8/9, the peak data rate can reach up to 100 Mbps and 50 Mbps within 20 MHz for the downlink and uplink, respectively [26]. With the advent of the CA strategy, Release 10 offers 3 Gbps for the downlink and 1.5 Gbps for the uplink in 100 MHz of spectrum. Finally, Release 13 enhanced the peak data rates to 25 Gbps in the downlink and 12.5 Gbps in the uplink, considering a bandwidth of 640 MHz.

2.1.5 Fifth-Generation Mobile Networks (5G)

Although the LTE is still under development, the industry is already defining the requisites of the next mobile generation network. The 5G network will focus not only on the mobile devices used by humans, but also will be concerned with Machine-Type Communications (MTC). This is why 5G is commonly related to the term *networked society* [27].

As occurred for the past generations, in 2013, the ITU-R initiated the activities to define the next IMT requirements (5G), referred to as IMT-2020.

In this sense, the main capabilities and requirements of the IMT-2020 are:

- **Data rates:** it is expected a peak data rate of 20 Gbps. The user experienced rate, i.e., the data rate in a large coverage for the majority of users, is expected to be 100 Mbps.
- **Latency:** the latency requirement was set to 1 ms.
- **Mobility:** maximum speed of 500 km/h.

- **Spectrum efficiency:** since 4G already presents a high spectrum efficiency, the requirement was set to 3 times the target for 4G networks.
- **Network energy efficiency:** the IMT-2020 network should not demand more energy than the IMT-Advanced, but should deliver an enhanced performance. In practice, this means that the network energy efficiency should be improved by a factor of 100.
- **Area traffic capacity:** it is related to the spectrum efficiency, bandwidth and the density of the network. IMT-2020 should improve this capability by 100, when compared to IMT-Advanced.
- **Connection density:** defines the number of devices in a particular area. This target is relevant for MTC scenarios and should be improved by a factor of 10.

Figure 2.4 sumarizes the IMT-2020 capabilities.

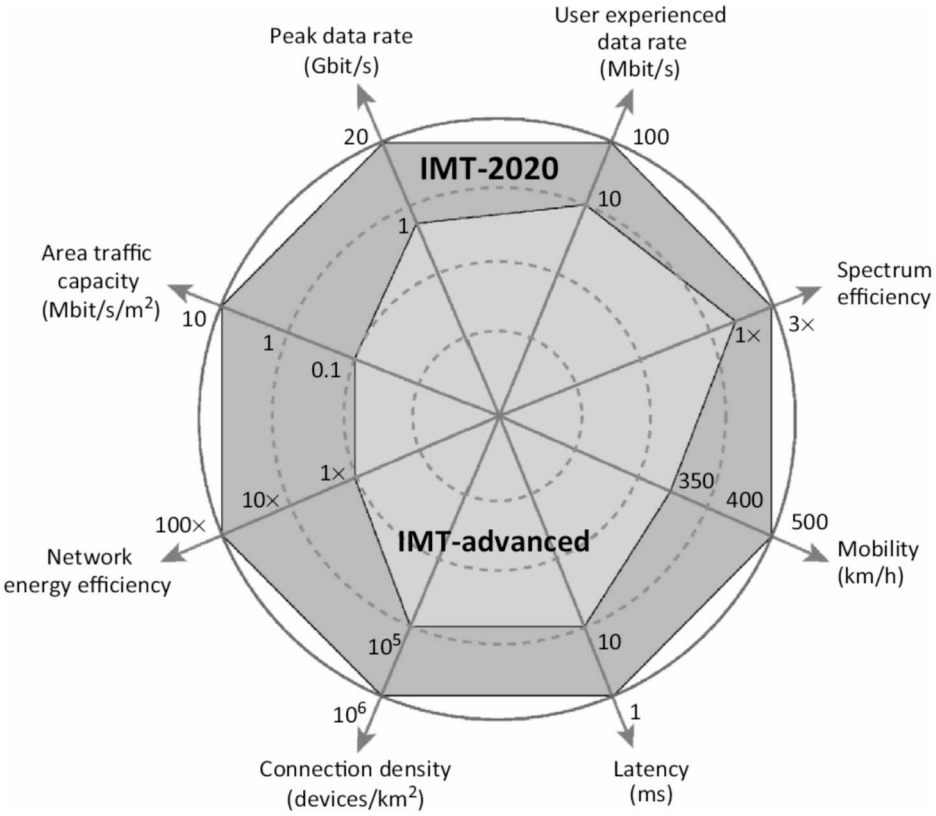


Figure 2.4: IMT-2020 capabilities (Adapted from [30]).

2.2 Network Architecture

The LTE system has been designed to have a flat architecture, which supports only the packet switched domain, i.e., it is an all-IP network. In this new network architecture, the core network is called System Architecture Evolution (SAE), while the term LTE covers the radio access network, the air interface and the mobile. Officially, the whole system is known as the Evolved Packet System (EPS), while the acronym LTE refers only to the evolution of the air interface [2]. However, despite of the official nomenclature, the name Long-Term Evolution is generally used to refer to the whole system.

At a high level, EPS has three main elements: the Evolved Packet Core (EPC), Evolved UMTS Terrestrial Radio Access Network (E-UTRAN), and the User Equipment (UE).

While the EPC consists of many logical nodes, the access network is made up of essentially just one node, the evolved NodeB (eNodeB), which connects to the UEs. Figure 2.5 shows the main elements of the EPC and the interfaces that connect the elements of the EPS. Next, we describe each one of these elements in more details.

2.2.1 User Equipment

The User Equipment (UE) can be divided in two parts: the Mobile Termination (MT), responsible for all communication functions, and the Terminal Equipment (TE), which terminates the data streams. For instance, the mobile termination might be a plug-in LTE card for a laptop and the terminal equipment would be the laptop itself [2].

The UE has a smart card called Universal Integrated Circuit Card (UICC). The UICC is also colloquially known as the SIM card. This card run an application known as Universal Subscriber Identity Module (USIM). This application stores user-specific data such as the user's phone number and home network identity. USIM is also used to identify and authenticate the user and to derive security keys for protecting the radio interface transmission [31].

The UE is the place where all the communication applications will run. It provides an interface between the end user and the system. It has also a wide variety of radio capabilities and features for mobility management, such as handover and terminal location reports.

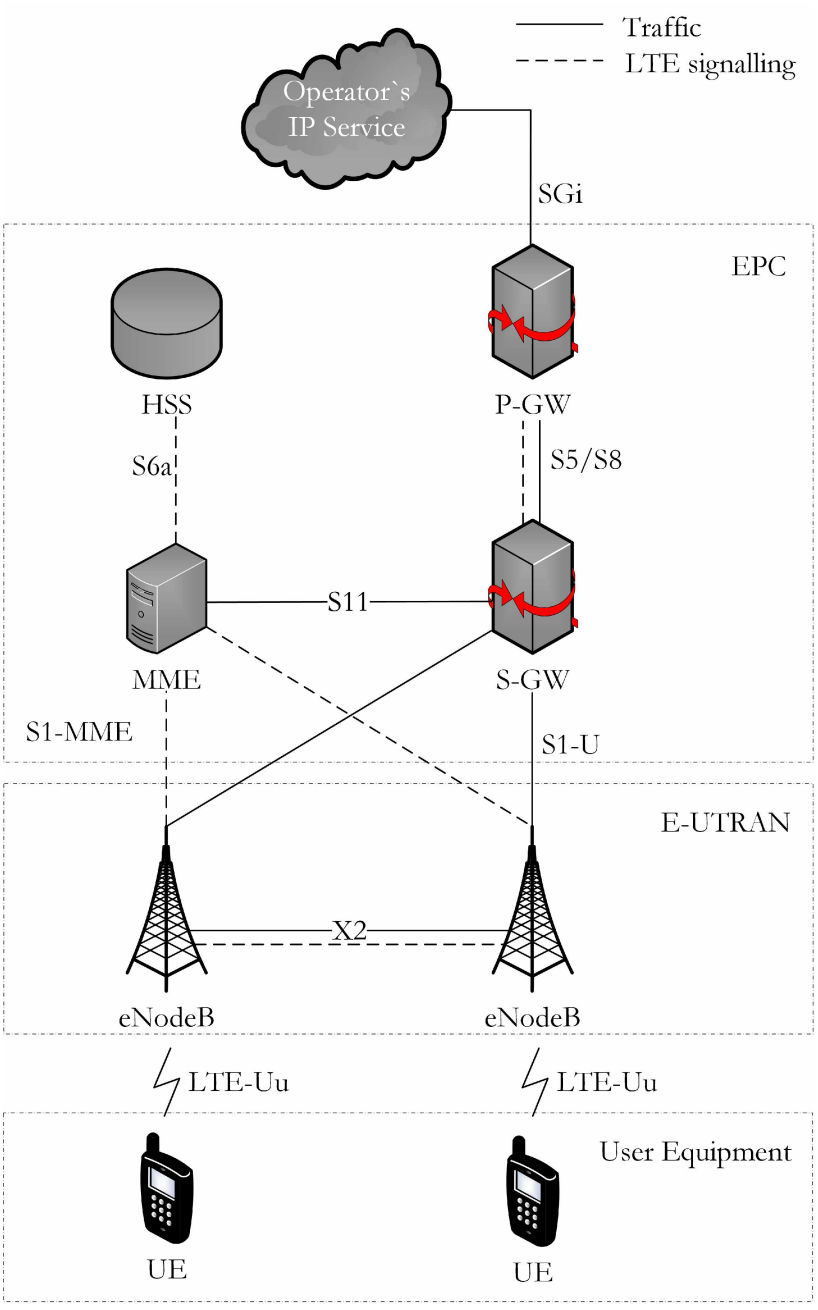


Figure 2.5: The EPS network elements.

2.2.2 E-UTRAN

The E-UTRAN is responsible for the radio communication between the UE and the EPC. The E-UTRAN is composed solely by the eNodeB. Each eNodeB controls a set of UEs distributed in one or more cells. On the other hand, a UE communicates with just one eNodeB and belongs to one cell at time[2].

[2]Release 12 introduced the Dual Connectivity concept. This feature allows a UE to be connected to two eNodeBs at the same time. This feature will not be considered in this thesis.

The eNodeB is responsible for the Radio Resource Management (RRM). This means that the eNodeB handles, for example, the radio resource allocation, providing the required Quality of Service (QoS). The eNodeB also handles the low-level operation of all its UEs, by sending them signaling messages related to those radio transmissions, such as handover commands [2].

As one can see in Figure 2.5, the eNodeB presents three interfaces. The X2 interface, between two eNodeBs, which is mainly used for signaling and packet forwarding during handover. The S1-U interface, between eNodeB and Serving Gateway (S-GW), which is used for user plane traffic. The S1-MME interface, between eNodeB and the Mobile Management Entity (MME), which is used for control plane traffic.

2.2.3 Evolved Packet Core

The EPC is the main component of the SAE. It consists of the following functional elements:

- **Serving Gateway (S-GW)**: all IP packets of users are transferred through the S-GW, which serves as the local mobility anchor for the data bearers when the UE moves between eNodeBs.
- **Packet Data Network (PDN) Gateway (P-GW)**: the P-GW is the EPC's point of contact with the outside world. Through the SGi interface, each PDN gateway exchanges data with any external devices or packet data networks.
- **Mobility Management Entity (MME)**: the MME is the key control element for the LTE access network. It is responsible for authenticating the user, UE tracking and paging procedure. It is also responsible for choosing the S-GW for UEs and it provides the control plane function for mobility between LTE and 2G/3G access networks.
- **Home Subscriber Server (HSS)**: the HSS contains users's SAE subscription data such as the EPS-subscribed QoS profile and any access restrictions for roaming. It also holds information about the PDNs to which the user can connect. In addition, the HSS holds dynamic information such as the identity of the MME to which the user is currently attached or registered.

There are other elements of the EPC that were not presented here. Interested readers can refer to [25] to find further information on this topic.

2.3 Protocol Architecture

Each one of the interfaces indicated in Figure 2.5 is associated with a protocol stack, which the network elements use to exchange data and signalling messages. The protocol architecture is divided in two groups:

- **user plane:** protocols used to handle data of interest of the user.
- **control plane:** protocols used to handle signalling messages.

Considering the air interface (Uu), we have the definition of two more terms: *Access Stratum* (AS) and *Non Access Stratum* (NAS). Access Stratum is a functional layer defined to allow the communication in the air interface between the UE and the eNB. In contrast, the Non Access Stratum is another functional layer defined to provide communication between the UE and the core network elements. In short, the distinction is that the AS is for explicit dialogue between the UE and the radio network, while the NAS is for dialogue between the UE and core network nodes that is passed transparently through the radio network.

2.3.1 User Plane Protocols

Figure 2.6 shows an overview of the user plane protocols.

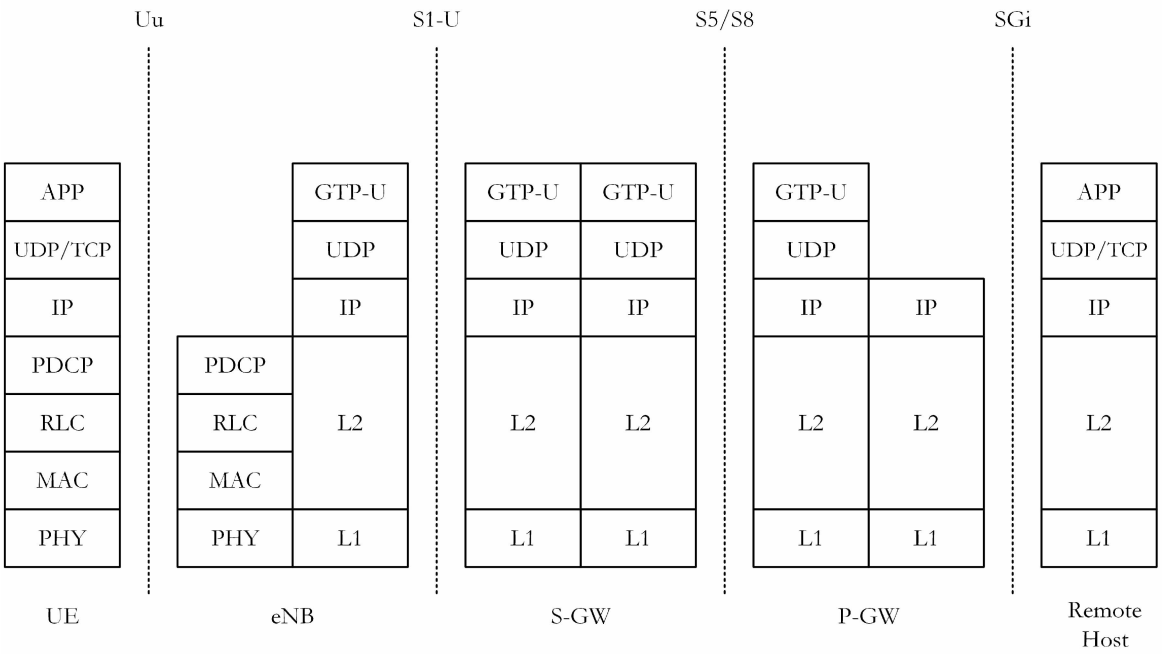


Figure 2.6: User plane protocol stacks (Adapted from [2]).

On the air interface Uu, we can identify three protocols at Layer 2:

- **Packet Data Convergence Protocol (PDCP):** is responsible for higher-level transport functions that are related to header compression and security.
- **Radio Link Control (RLC):** the main functions of the RLC layer are segmentation and reassembly of upper layer packets. In some configurations, this layer is also responsible for ensuring reliable delivery for data streams that need to arrive correctly.
- **Medium Access Control (MAC):** this layer performs multiplexing of data from different radio bearers and is responsible for scheduling data transmissions between the mobile and the base station.

In the S1-U and S5/S8 interfaces the GPRS Tunneling Protocol User Part (GTP-U) handles the data flow between the pairs eNB/S-GW and S-GW/P-GW.

2.3.2 Control Plane Protocols

Figure 2.7 shows an overview of the control plane protocols.

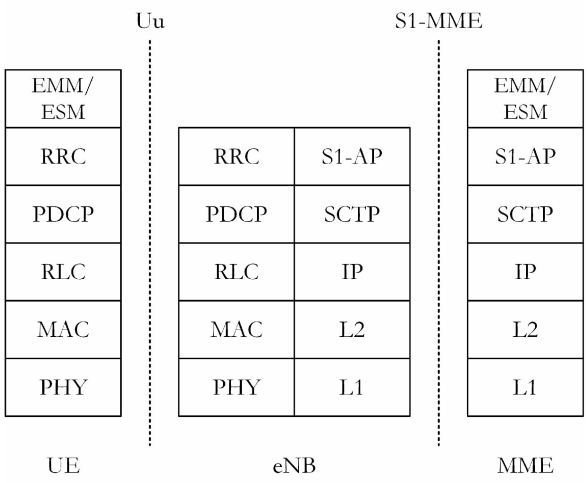


Figure 2.7: Control plane protocol stacks (Adapted from [2]).

On the air interface Uu, the control messages are handled by the Radio Resource Control (RRC) protocol. In the S1 interface, the MME controls the base stations using the S1 Application Protocol (S1-AP). The Stream Control Transmission Protocol (SCTP) is a TCP-based protocol that includes extra features that make it more suitable for the delivery of signalling messages.

In the NAS logical layer, the MME controls a mobile's high-level behavior using two protocols: the EPS Session Management (ESM), which controls the data streams

through which a mobile communicates with the outside world, and EPS Mobility Management (EMM), which handles internal bookkeeping within the EPC. The network transports EMM and ESM messages by embedding them into lower-level RRC and S1-AP messages and then by using the transport mechanisms of the Uu and S1 interfaces [2].

2.4 Logical, Transport and Physical Channels

The information that flows between the different protocols are known as *channels*. Figure 2.8 shows the air interface protocol stack from the UE point of view. This figure also indicates the location of the different channels between the protocols. LTE uses several different types of logical, transport and physical channels, which are distinguished by the kind of information they carry and by the way in which the information is processed. The physical layer can be divided in three parts: (i) the *transport channel processor* is responsible for the error management procedures; (ii) the *physical channel processor* applies the techniques of OFDMA, SC-FDMA and multiple antenna transmission; (iii) finally, the analogue processor converts the signal to be transmitted. In the next sections, we offer more details of how these channels work.

2.4.1 Logical Channels

Table 2.1 lists the logical channels. One can see that there are channels reserved for user and control plane purposes. The most important logical channels are the Dedicated Traffic Channel (DTCH), which carries data to or from a single mobile, and the Dedicated Control Channel (DCCH), which carriers most of signalling messages.

Table 2.1: LTE logical channels.

Channel	Name	Information carried
DTCH	Dedicated Traffic Channel	User plane data
DCCH	Dedicated Control Channel	Signalling
CCCH	Common Control Channel	Signalling
PCCH	Paging Control Channel	Paging messages
BCCH	Broadcast Control Channel	System information

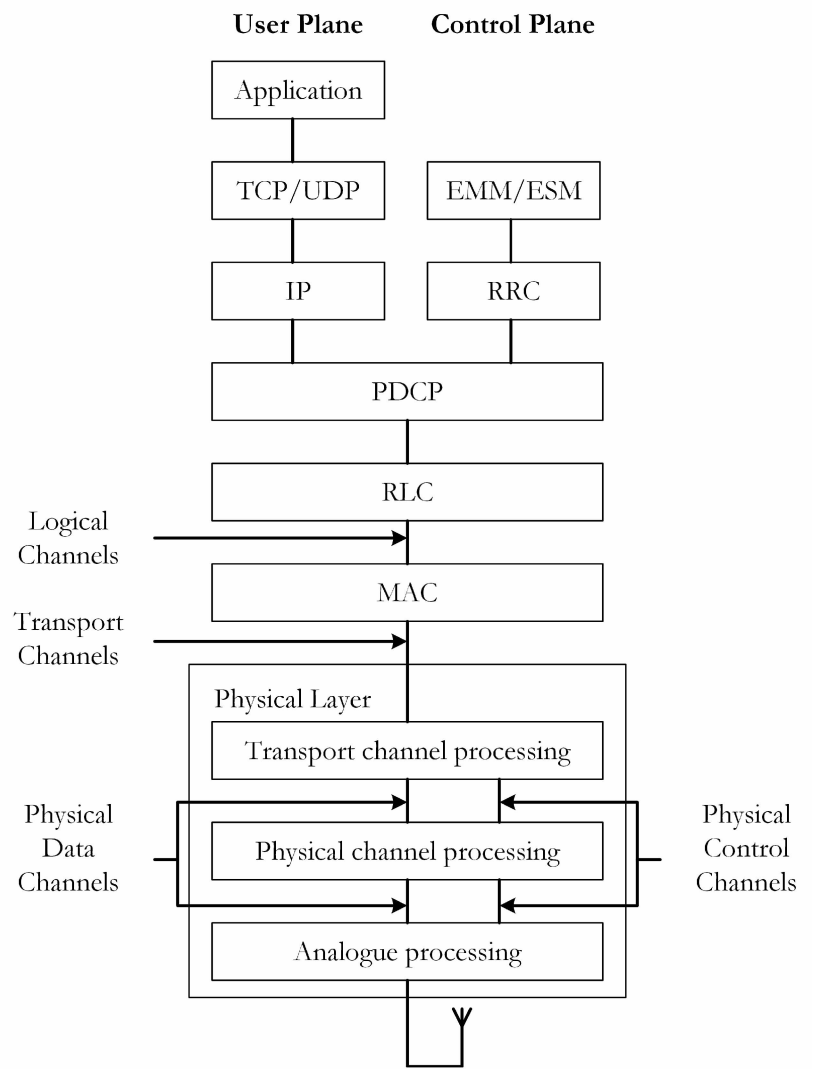


Figure 2.8: Air interface protocol stack (Adapted from [2]).

2.4.2 Transport Channels

The transport channels are listed in Table 2.2. The most important transport channels are the Downlink Shared Channel (DL-SCH) and the Uplink Shared Channel (UL-SCH), which carry the majority of data and signalling messages in the air interface. These two channels are the only transport channels allowed to adapt their coding rate according to the channel quality. They can also make use of techniques of automatic repeat request to recover from transmission losses.

The Random Access Channel (RACH) is a special channel that the mobile uses to contact the base station without any prior scheduling. It is used by the UE to get connected to the eNodeB.

Table 2.2: LTE transport channels.

Channel	Name	Information carried
DL-SCH	Downlink Shared Channel	Downlink data and signalling
UL-SCH	Uplink Shared Channel	Uplink data and signalling
RACH	Random Access Channel	Random Access Request
PCH	Paging Channel	Paging messages
BCH	Broadcast Channel	Master information block
MCH	Multicast Channel	MBMS data

2.4.3 Physical Channels

Table 2.3 lists the LTE physical channels. The most important physical channels are the Physical Downlink Shared Channel (PDSCH) and the Physical Uplink Shared Channel (PUSCH). PDSCH carries data from DL-SCH and PCH. PUSCH is responsible for transporting data from UL-SCH. These two channels are the only physical channels allowed to change their modulation schemes in response to channel quality variations.

Table 2.3: LTE physical channels.

Channel	Name	Information carried
PDSCH	Physical Downlink Shared Channel	DL-SCH and PCH
PUSCH	Physical Uplink Shared Channel	UL-SCH
PRACH	Physical Random Access Channel	RACH
PBCH	Physical Broadcast Channel	BCH
PMCH	Physical Multicast Channel	MCH
PDCCH	Physical Downlink Control Channel	DCI
PUCCH	Physical Uplink Control Channel	UCI
PCFICH	Physical Control Format Indicator Channel	CFI
PHICH	Physical Hybrid ARQ Indicator Channel	HI

Table 2.3 also lists the physical control channels. These channels are related to specific control information. Table 2.4 enumerates the types of control information used in LTE.

The Downlink Control Information (DCI) is used by the eNodeB to alert mobiles of upcoming downlink data or to notify the mobile when it can transmit in the uplink shared channel. DCI is also used to adjust the transmission power of the UEs.

Table 2.4: LTE control information.

Channel	Name	Information carried
DCI	Downlink Control Information	Downlink scheduling commands Uplink scheduling grants Uplink power control commands
UCI	Uplink Control Information	Hybrid ARQ acknowledgements Channel Quality Indicator (CQI) Pre-coding Matrix Indicators (PMI) Rank Indications (RI) Scheduling Requests (SR)
CFI	Control Format Indicator	Size of downlink control region
HI	Hybrid ARQ Indicator	Hybrid ARQ acknowledgements

The Uplink Control Information (UCI) has several fields. HARQ ACKs (Hybrid ARQ Acknowledgements) are used to acknowledge the packets received on the DL-SCH. The Channel Quality Indicator (CQI) describes the quality of the connection between the UE and the base station. The Pre-coding Matrix Indicators (PMI) and Rank Indications (RI) are used to support multiple antenna techniques. Scheduling Requests (SR) are sent by the mobile, asking for resources to transmit on the PUSCH.

Control Format Indicators (CFI) describe the data organization in the downlink. HARQ Indicators (HI) are used to acknowledge packets received on the UL-SCH.

2.5 Summary

In this chapter, we took an overview of the main terms and concepts that describe the LTE network. One can note that LTE is a very complex network, with several elements and protocols. From the network elements presented in this chapter, we are more interested in the communication between the UE and the eNB. In the next chapter, we will study how the eNB allocates its resources among the users attached to it.

"Technology is a useful servant but a dangerous master."

- Christian Lous Lange

3

Resource Allocation in the LTE Network

COMMUNICATION NETWORKS are designed to serve a particular number of users. However, the network resources are always limited, and they have to be split among the users. This process is called resource allocation. In this chapter, we present a brief description of how the LTE network shares its resources among the UEs.

3.1 Physical Layer Design

Multiple access techniques allow the base station to communicate with several mobile terminals simultaneously. As we said in Section 2.1.4, LTE uses OFDMA and SC-FDMA as the techniques for multiple access in downlink and uplink, respectively. In the next sections, we describe how these techniques work and their main differences.

3.1.1 OFDMA

In order to allow the eNodeB to serve multiples UEs, OFDMA performs the transmissions at different times and frequencies. These set of frequencies are known as *subcarriers*. Each subcarrier is modulated with a conventional modulation scheme like

Quadrature Phase Shift Keying (QPSK) or 16-QAM, for example. The use of orthogonal subcarriers is also important to mitigate the Inter Symbol Interference (ISI). Figure 3.1 shows how the subcarriers are spaced in a manner that the interference among them is decreased. At the instant the amplitude of a particular subcarrier is sampled, the amplitude of the other subcarriers equals zero. In LTE, the subcarrier spacing (Δf) is equal to 15 kHz.

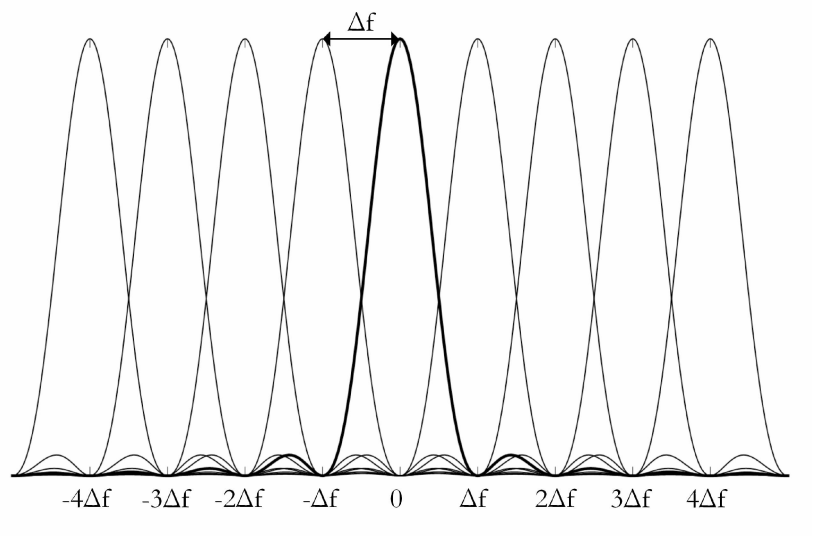


Figure 3.1: OFDMA subcarrier spacing.

OFDMA is efficiently implemented via Fast Fourier Transform (FFT) and also present other advantages like high spectral efficiency and bandwidth scalability [4].

Figure 3.2 illustrates a block diagram of the OFDMA transmitter used by the eNodeB. In OFDMA transmission, each symbol is mapped to a particular subcarrier. Therefore, each subcarrier only carries information related to one specific symbol.

OFDMA works well in the downlink. However, despite of the set of advantages, OFDMA has one disadvantage: it presents a large variation in the power of the transmitted signal. These variations require the use of large and expensive power amplifiers, which can deal with the power variations without reaching the saturation region. The base station can afford this kind of power amplifiers. However, the same is not true for the UEs. The mobile transmitter must be cheap, small and energy efficient. This fact makes OFDMA unsuitable for the uplink.

3.1.2 SC-FDMA

The power variations described in the last section occur because there is a one-to-one mapping between symbols and subcarriers. In this sense, if we mix the symbols before

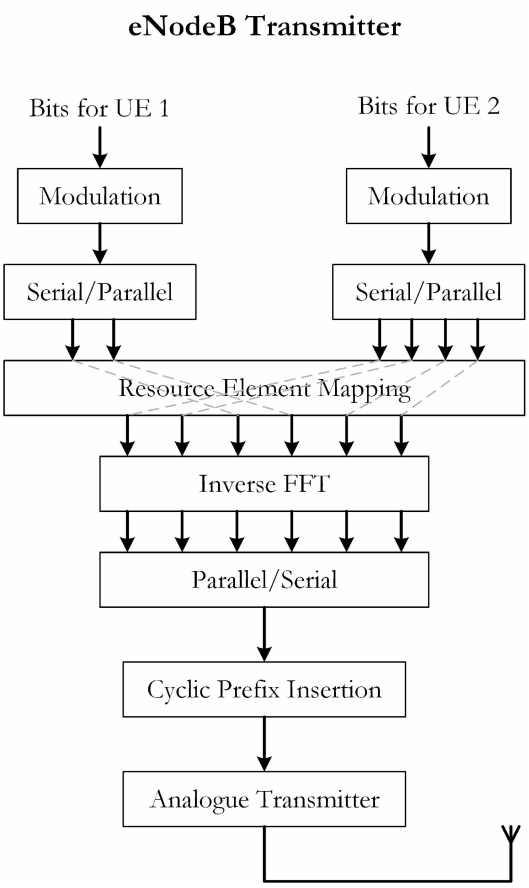


Figure 3.2: Block Diagram of the OFDMA transmitter (Adapted from [2]).

placing them into the subcarriers, we might be able to adjust the transmitted signal and reduce its power variation.

One of the most suitable mixing operation is a forward FFT. In this sense, SC-FDMA uses a forward FFT to mix the symbols and avoid large power variations. The process is depicted in Figure 3.3.

Besides the addition of the forward FFT before the resource element mapping, we also can identify the following difference between OFDMA and SC-FDMA: the first difference arises because the technique is used on the uplink. Because of this, the mobile transmitter uses only some of the subcarriers. The others are set to zero, and are available for the other mobiles in the cell. The second difference is given by the fact that each mobile transmits using a single, contiguous block of subcarriers, without any internal gaps. This is necessary to keep the power variations as lowest as possible [2].

An alternative description is provided in [32] and depicted in Figure 3.4. This figure shows, in frequency and time, how OFDMA and SC-FDMA would each transmit a sequence of 8 QPSK data symbols.

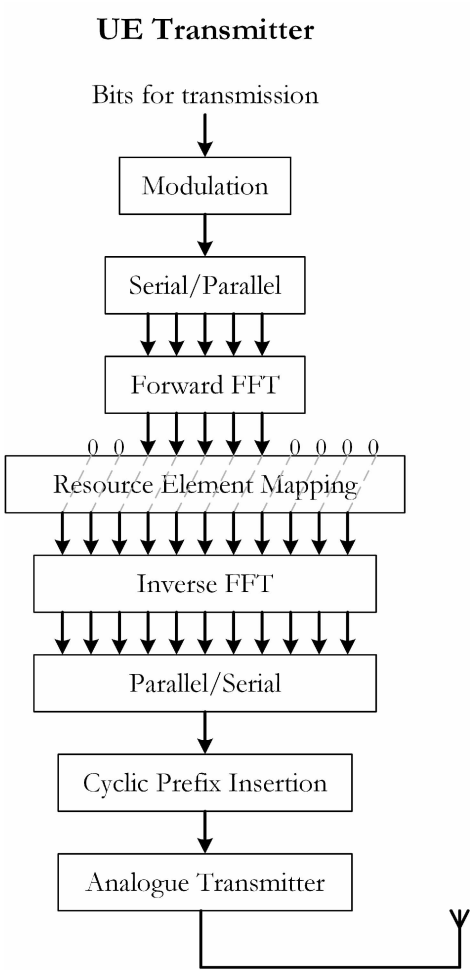


Figure 3.3: Block Diagram of the SC-FDMA transmitter (Adapted from [2]).

In the example, there are four subcarriers. For OFDMA, four symbols are used in parallel to modulate their own subcarrier. Each data symbol occupies 15 kHz for the period of one OFDMA symbol (66.7 μs). The parallel transmissions allow the data symbols be the same length as the OFDMA symbols.

In the SC-FDMA case, the data symbols are transmitted sequentially. In this sense, four data symbols are transmitted sequentially in one SC-FDMA symbol period. The SC-FDMA symbol period is the same length as the OFDMA symbol. However, because of sequential transmission, the data symbols are shorter being $66.7/M\mu$s, where M is the number of subcarriers. This phenomenon requires more bandwidth, so each data symbol occupies 60 kHz of spectrum rather than the 15 kHz used in OFDMA [32].

Since the RB contiguity constraint can be very challenging for the scheduling process, Release 10 introduced a feature that allows a mobile to use a non contiguous allocation of sub-carriers. This is possible by using a Discret Fourier Transform Spread Orthogonal Frequency Division Multiple Access (DFTS-OFDMA) [2]. However, to be

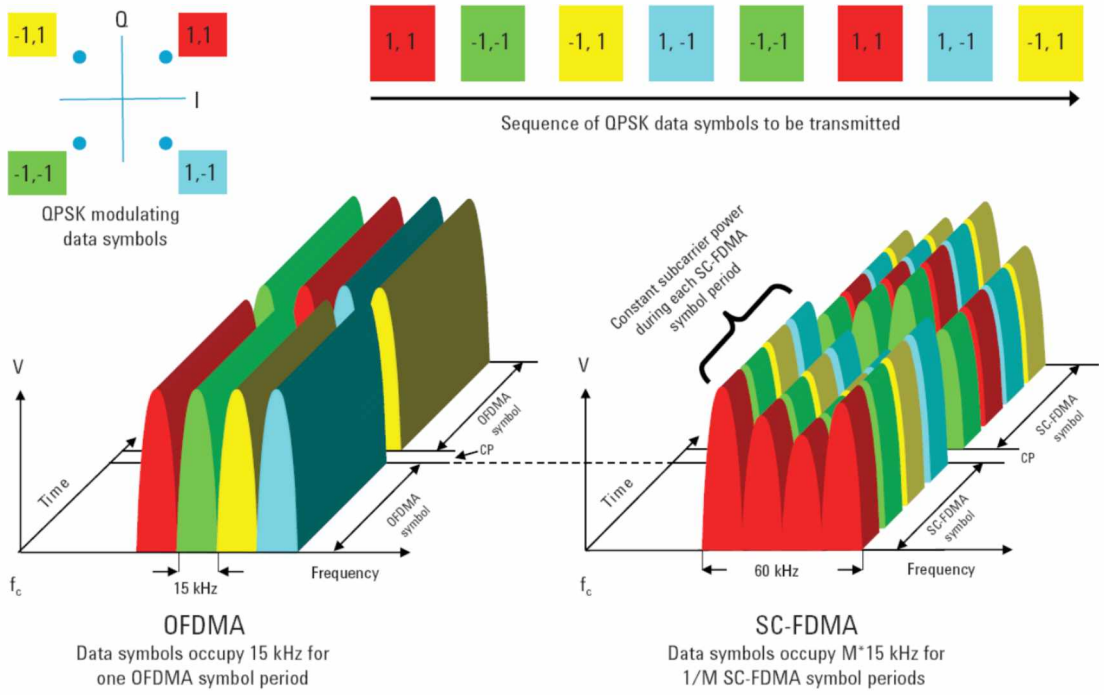


Figure 3.4: Comparison of OFDMA and SC-FDMA transmitting a series of QPSK data symbols [32].

more specifically, in Release 10 the uplink resource assignment consist of a maximum of two frequency-separated groups of RBs [27]. Then, the contiguity constraint can be broken, but only for two clusters of resource blocks. In practice, this gives more flexibility to the scheduling algorithm, but the constraint still need to be considered.

3.2 Cyclic Prefix

Besides the subcarrier spacing, LTE also makes use of the *cyclic prefix insertion* to minimize the effects of the interference in the reception of the symbol. The cyclic prefix insertion consists of a technique in which an ending piece of the symbol is copied to its front. Figure 3.5 shows the cyclic prefix insertion operation.

3.3 The Resource Grid Structure

In the last sections, we briefly presented how LTE allows multiple users to have access to the base station resources simultaneously. We also introduced the concept of the symbols as the minor unit of data in the air interface. In this section, we will study how

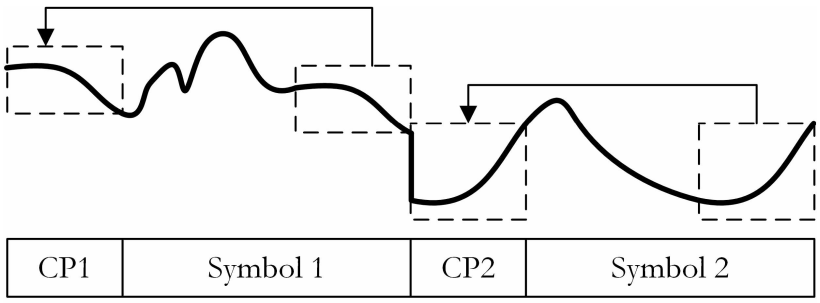


Figure 3.5: Operation of cyclic prefix insertion (Adapted from [2]).

the base station resources are grouped and organized to optimize the allocation process.

The symbols are grouped into *slots*, whose duration is 0.5 ms. This grouping can be performed in two ways [2]:

- **normal cyclic prefix**: each symbol is preceded by a cyclic prefix that is usually 4.7 μs long. The first cyclic prefix is set to 5.2 μs, since it is not possible to fit seven symbols of the same size into a slot.
- **extended cyclic prefix**: each symbol is preceded by a cyclic prefix of 16.7 μs. In this configuration, the number of symbols in the slot is reduced to six.

Using the normal cyclic prefix the receiver can deal with ISI presenting a delay spread of 4.7 μs, corresponding to a path difference of 1.4 km between the lengths of the longest and shortest rays. This is normally enough. However, for unusually larger cells, the extended cyclic prefix may be more appropriated, since it supports a maximum path difference of 5 km and delay spread of 16.7 μs. Figure 3.6 shows the organization of the symbols into slots using these two approaches.

At a higher level, slots are grouped into *sub-frames*. In FDD mode, two slots form a sub-frame of 1 ms. Sub-frames are defined for scheduling purposes. When the eNodeB transmit to a UE, it schedules its transmissions one sub-frame at time. The sub-frame is also known as the Transmission Time Interval (TTI). In the uplink, the process is similar.

A sequence of 10 sub-frames form a *frame* of 10 ms. Figure 3.7 depicts the LTE *frame structure type 1*, used in the FDD mode. There is also a *frame structure type 2*, used in TDD mode. In this thesis, we will work only with FDD mode. Interested readers can get more information about frame structure type 2 in [2].

So far, we have seen how LTE uses multiple subcarriers in the frequency domain and the concept of slots in the time domain. Next, we explore how these two domains work together to form the *resource grid*. Figure 3.8 illustrates the LTE resource grid,

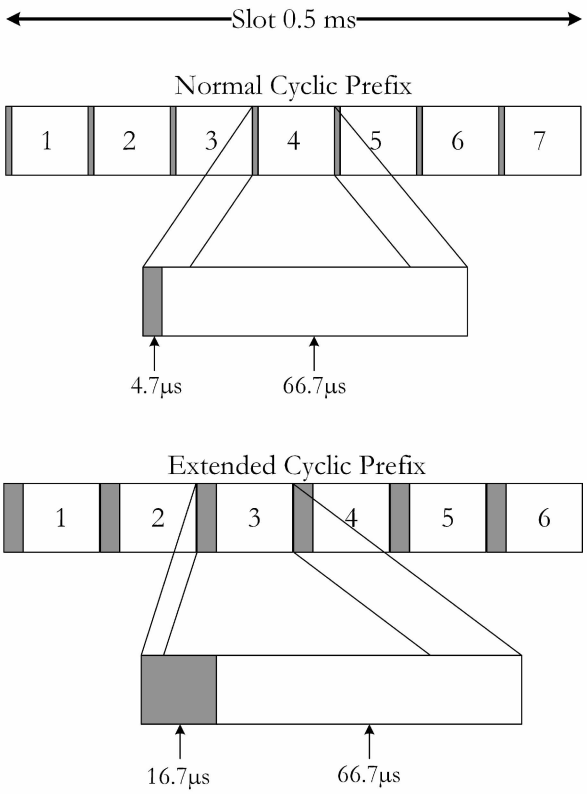


Figure 3.6: Organization of symbols into slots using the normal and extended cyclic prefixes.

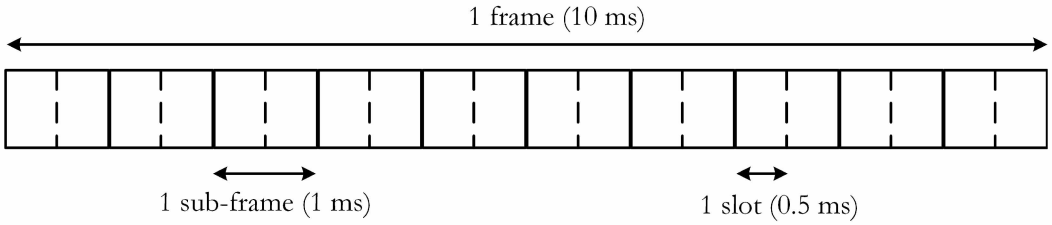


Figure 3.7: Frame structure type 1, used in FDD mode.

considering the normal cyclic prefix. The basic unit is the *resource element*, which is formed by one symbol in time and one subcarrier in frequency. Resource elements are grouped into *Resource Blocks* (RB). Each RB spans one slot (0.5 ms) by 180 kHz (12 subcarriers). In this sense, the RB is the unit used by the base station to allocate its resources among the mobiles.

3.4 Bandwidth

An LTE cell can be configured with several different bandwidths, according to Table 3.1. This set of bandwidths offers to the network operators a very flexible environment

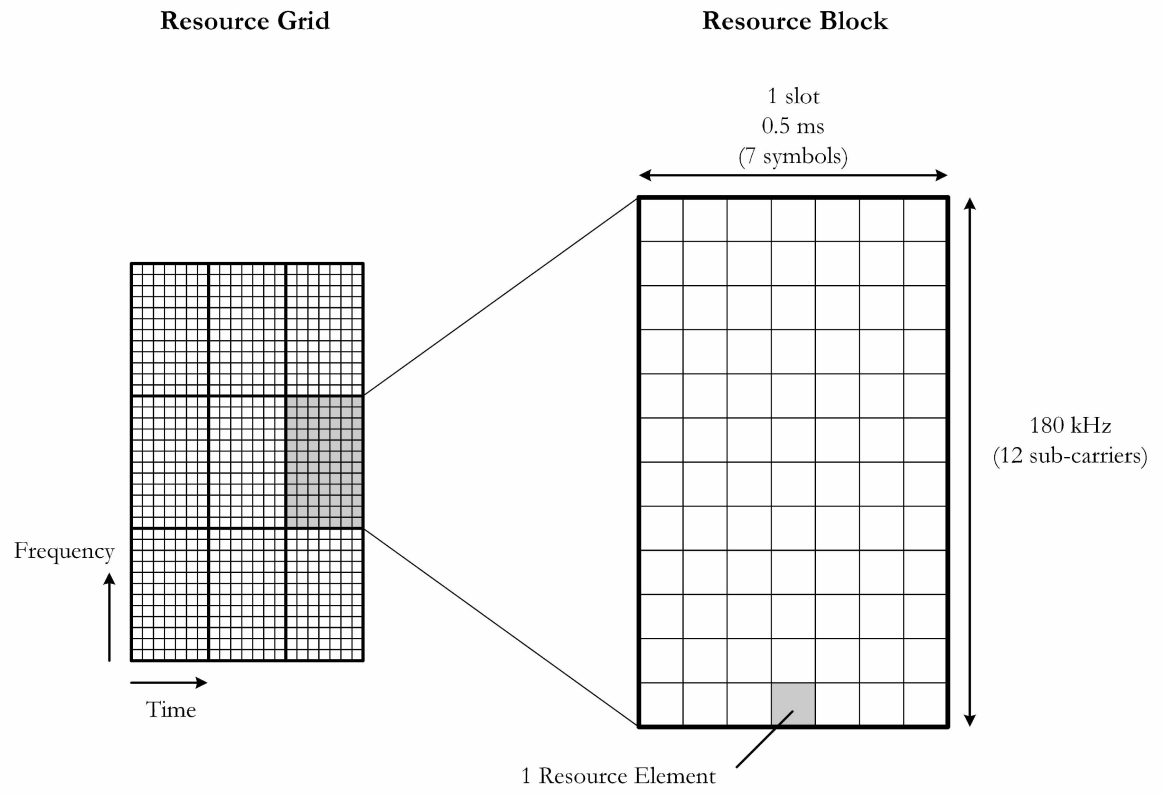


Figure 3.8: LTE resource grid considering the normal cyclic prefix (Adapted from [2]).

for the deployment of LTE. For example, 1.4 MHz is close to the bandwidth earlier used by cdma2000, 5MHz is the same bandwidth used by WCDMA, while 20 MHz allows an LTE base station to operate at its highest possible data rate[1] [2].

Table 3.1: LTE supported bandwidths.

Bandwidth	Number of RBs	Number of subcarriers
1.4 MHz	6	72
3 MHz	15	180
5 MHz	25	300
10 MHz	50	600
15 MHz	75	900
20 MHz	100	1200

[1]Not considering the Carrier Aggregation technique.

3.5 Link Adaptation

Link Adaptation (LA) is the name given to the techniques that allow LTE to dynamically adjust the transmitted information data rate (modulation scheme and channel coding rate) to match the current radio channel capacity for each user. Link Adaptation is generally based on Adaptive Modulation and Coding (AMC) techniques.

The AMC degrees of freedom are related to the available modulation and coding schemes [25]:

- **Modulation Scheme**: if the interference levels are high, it is necessary to apply a low-order modulation like BPSK or QPSK(used for PUCCH modulation), which is more robust, but provides a lower transmission bit rate. On the other hand, when the SINR is sufficiently high, it's better to use a high-order modulation (16QAM, 64QAM), which offers a higher bit rate.
- **Code rate**: the same analysis is true for the code rate. Depending on the radio link conditions, a lower code rate can be used in poor channel conditions and a higher code rate for high SINR.

In LTE, all Resource Blocks allocated to one user in a sub-frame must use the same Modulation and Coding Scheme (MCS).

An important input to select the most suitable modulation and coding rate in the downlink is the Channel Quality Indicator (CQI) feedback, sent by the UE in the uplink. CQI feedback is an index that takes into account the SINR and the characteristics of the UE's receiver. The UE reports the highest MCS that it can decode with a Block Error Rate (BLER) probability not exceeding 10%.

In the time domain, the LTE standard defines two types of CQI report: *periodic* and *aperiodic*. The UE can be configured to send periodic CQI report to the eNodeB using the PUCCH. This channel is exclusive for this kind of report. On the other hand, the eNodeB can request a CQI report from the UE. In this case, we have an aperiodic CQI, which is sent through the PUSCH. The CQI report is embedded into a resource which is scheduled for uplink transmission [25].

There is also different granularity for CQI reporting process. CQI reports can be classified as [25]:

- **Wideband feedback**: the UE reports one wideband CQI value for the whole system bandwidth.
- **eNodeB-configured sub-band feedback**: the UE reports a CQI value for each

sub-band and a wideband CQI value for the whole system bandwidth.

- **UE-selected sub-band feedback**: the UE reports a CQI value for a set of preferred sub-bands and a wideband CQI for the whole system bandwidth.

For the uplink, the link adaptation process is similar. The selection of modulation and coding schemes is under the control of the eNodeB. The modulation can be selected among the QPSK, 16QAM and 64QAM schemes. The main difference lies on the fact that there is not a CQI feedback. The eNodeB can directly make its own estimate of the supportable uplink data rate by channel sounding, using the Sounding Reference Signals (SRS), for example. In this sense, the SRS are one of the mechanisms of Channel State Information (CSI) report.

If the UE has received a grant for uplink transmission, it sends the CSI data using the SRS in the PUSCH. For the case when the UE has not received a grant for uplink transmission, it uses the PUCCH to send the channel quality information. There is also another case, when the UE does not have access to a PUCCH. Then, the UE must use the PRACH to get access to a PUCCH.

3.6 HARQ

Hybrid Automatic Repeat Request (HARQ) is an error management technique based on retransmissions. The HARQ entity is part of the MAC layer and is responsible for handling HARQ operations. These operations include transmission and retransmission of packets, and reception and processing of ACK/NACK signalling. The transmission operations are based on the Stop-And-Wait (SAW) approach. SAW operation means that upon transmission of a Transport Block (TB), the transmitter stops further transmissions and waits the feedback from the receiver. When a NACK is received, or when a certain time elapses without receiving any feedback, the transmitter retransmits the packet.

Unfortunately, this technique cannot use the transmission resources during the period between the first transmission and the retransmission. This is not efficient. To improve the efficiency of SAW operation, LTE uses up to eight independent processes that work in parallel mode, so that all the transmission resources can be used [25].

Generally, HARQ schemes can be categorized as either *synchronous* or *asynchronous*, and each scheme can be either *adaptive* or *non-adaptive*.

In a synchronous HARQ scheme, retransmissions occur at predefined times relative to the initial transmission. In this scheme, the signalling information can be inferred

from the transmission timing. On the other hand, in an asynchronous HARQ scheme, the retransmissions can occur at any time and signalling information is necessary.

For adaptive HARQ scheme, transmission attributes such as the MCS can be changed at each retransmission, according to the variations of the radio channel conditions. In a non-adaptive HARQ scheme, retransmissions are performed by using the same transmission attributes as those of the previous transmission, or by changing the attributes according to a predefined rule.

In summary, synchronous and non-adaptive HARQ schemes reduce the signalling overhead while asynchronous and adaptive HARQ schemes allow more flexibility in scheduling.

In LTE, downlink is based on asynchronous adaptive HARQ, while in the uplink is used synchronous HARQ scheme and the retransmission can be either adaptive or non-adaptive [25].

3.7 Buffer Status Report

The UE uses Buffer Status Report (BSR) messages to inform the base station about how much data it has available for transmission. These messages are important in the process of provision of QoS, since the information is used by the scheduler to allocate the resources.

There are three types of BSR [2]:

- **Regular BSR**: a mobile sends this report in three situations. The first one occurs if data become ready for transmission, when the transmission buffers were previously empty. The second situation is triggered when data become ready for transmission on a logical channel with a higher priority than the current buffers. Finally, this report is also sent if a timer expires while data are waiting for transmission. The mobile expects the base station to reply with a scheduling grant.
- **Periodic BSR**: this report is sent at regular intervals during data transmission on the PUSCH.
- **Padding BSR**: this report is triggered if the UE has enough spare room during a normal PUSCH transmission.

BSR are reported on a per Radio Bearer Group (RBG) basis as a result of a compromise between the need of differentiation of data flows based on QoS requirements

and the signalling overhead. The number of RBG is limited to 4. Each RBG groups radio bearers with similar QoS requirements [4].

3.8 Power Headroom Report

The Power Headroom Report (PHR) is the difference between the UE's maximum transmit power and the estimated power for PUSCH transmission. The base station uses this information to support its uplink scheduling procedure, typically by limiting the data rate at which it asks the mobile to transmit [2].

A number of criteria are defined to trigger a PHR. These include:

- a significant change in estimated path-loss since the last PHR;
- the predefined timer expires.

3.9 Quality of Service Mechanisms

In a typical case, we have multiple applications running in the cells under control of a given eNodeB. There is also the case when we have multiple applications running in a particular UE at the same time. In both cases, these applications have different requirements to work, which we generally call Quality of Service (QoS) requirements. For example, a user starts a download using File Transfer Protocol (FTP). During this download, he decides to start a VoIP call. VoIP has more stringent QoS requirements in terms of delay and jitter than the FTP application, while the latter requires a much lower packet loss rate. In this sense, to support different QoS requisites, LTE has defined different *bearers*, each being associated with a QoS requirement [25].

The simplest definition for a bearer is getting it as a data pipe, which transfers information between the UE and the PDN gateway with a specific QoS (delay, data and error rates). The most important bearer is the EPS bearer [2]. Figure 3.9 shows the EPS bearer service architecture.

From Figure 3.9 one can see that the EPS bearer has to cross multiple interfaces. Therefore, the EPS bearer is broken down into three lower level bearers, namely, the radio bearer, the S1 bearer and the S5/S8 bearer. The information carried by the EPS bearer is grouped in one or more *service data flows*. These flows carry packets for a particular service such as video streaming application. In turn, each service data flow comprises one or more *packet flows*, such as the audio and video streams, which form

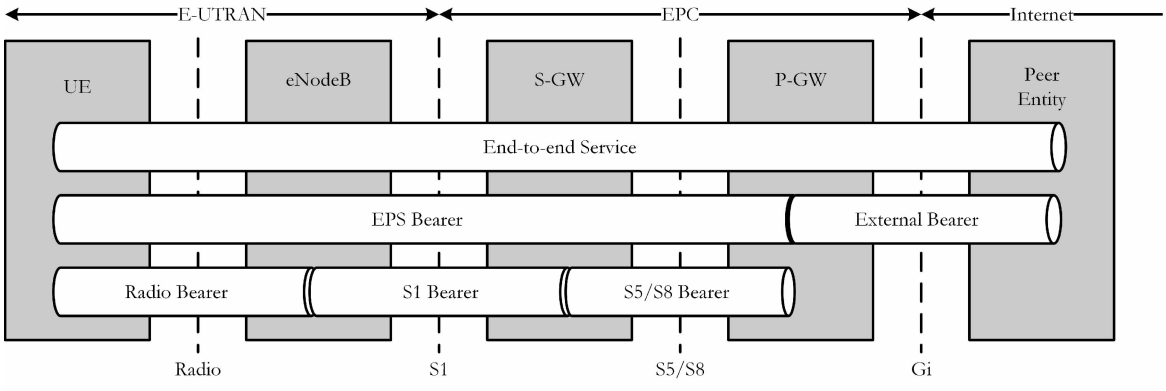


Figure 3.9: The overall EPS bearer service architecture [25].

that service. LTE provides the same quality of service to all the packet flows within a particular EPS bearer [2].

In order to provide QoS, 3GPP has defined a set of QoS parameters. They are [31] [2]:

- **QoS Class Identifier (QCI)**: it is an index that identifies a set of locally configured values for three QoS attributes: Priority, Delay and Loss Rate. QCI is signaled instead of the values of these parameters. Ten pre-configured classes have been specified in two categories of bearers, Guaranteed Bit Rate (GBR) and Non-Guaranteed Bit-Rate (Non-GBR) bearers.
- **QCI priority level**: helps the scheduling process. Low numbers are associated with a higher priority. Then, a congested network meets the packet delay budget of bearers with priority N, before moving on to bearers with priority $N + 1$.
- **Packet delay budget**: is an upper bound, with 98% confidence, for the delay that a packet receives between the UE and the P-GW.
- **Packet error/loss rate**: is an upper bound for the proportion of packets that are lost due to errors in transmission and reception.
- **Guaranteed Bit Rate (GBR)**: identifies the bit rate that will be guaranteed to the bearer.
- **Maximum Bit Rate (MBR)**: identifies the maximum bit rate for the bearer. Note that a Release 8 network is not required to support differentiation between the MBR and GBR, and the MBR value is always set to the GBR value.
- **Aggregate Maximum Bit Rate (AMBR)**: Non-GBR bearers are collectively associated with the per UE aggregate maximum bit rate.

- **Allocation and Retention Priority (ARP)**: indicates the priority of the bearer compared to other bearers. This provides the basis for admission control in bearer set-up, and further in a congestion situation if bearers need to be dropped.

Table 3.2 summarizes the standardized QoS Class Identifiers for LTE.

Table 3.2: Standardized QoS Class Identifiers (QCIs) for LTE [33].

QCI	Resource type	Priority	Delay budget (ms)	Loss rate	Example applications
1	GBR	2	100	10^{-2}	VoIP
2	GBR	4	150	10^{-3}	Video Call
3	GBR	3	50	10^{-3}	Real time gaming
4	GBR	5	300	10^{-6}	Streaming
5	Non-GBR	1	100	10^{-6}	IMS signaling
6	Non-GBR	6	300	10^{-6}	Application with TCP: browsing, email, FTP, etc.
7	Non-GBR	7	100	10^{-3}	Interactive gaming
8	Non-GBR	8	300	10^{-6}	Application with TCP: browsing, email, FTP, etc.
9	Non-GBR	9			

When a UE connects to a packet data network, the EPC sets up an EPS bearer, known as a *default bearer*. Then, the UE receives an IP address to use when communicating with that network. The UE can establish connections with other packet data networks and it receives an additional default bearer for every network that it connects to, together with an additional IP address.

After the establishment of a default bearer, a UE can also receive one or more *dedicated bearers*. This does not lead to the allocation of any new IP addresses, instead, each dedicated bearer shares an IP address with its parent default bearer. A dedicated bearer typically has a better quality of service than the default bearer can provide and particularly can have a guaranteed bit rate [2].

As we said before, the EPS bearer is divided into three lower level bearers. In this context, the GTP-U protocol creates bi-directional tunnels between the S1 and S5/S8 bearers and between the S5/S8 and external bearers. In turn, each tunnel is associated with two Tunnel Endpoint Identifiers (TEIDs), one for the uplink and one for the downlink.

As an example, let us consider a flow of data packets on the downlink. In Figure

3.10, a UE has two EPS bearers, which are carrying video and email packets that require different requirements of QoS. Remote hosts send the packets through the Internet and they finally reach the P-GW.

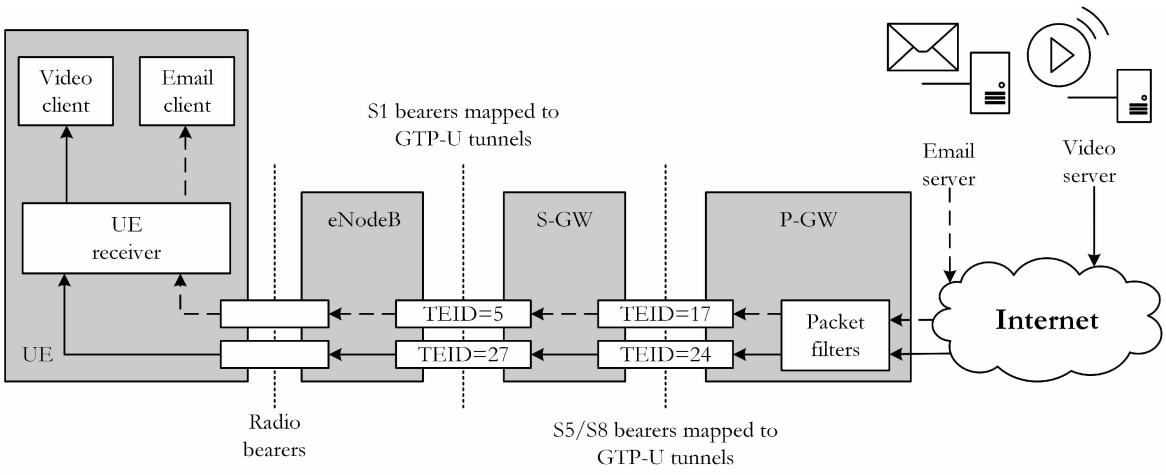


Figure 3.10: Downlink GTP tunneling [2].

The P-GW now has to assign each incoming packet to the correct EPS bearer. In order to achieve this, each EPS bearer is associated with a Traffic Flow Template (TFT). A TFT consists of a set of *packet filters*, one for each of the packet flows that make up the bearer. In turn, each packet filter has information about the IP addresses of the source and destination devices, and the UDP or TCP port numbers of the source and destination applications. Using packet filters, the P-GW inspects every incoming packet and assigns them to the correct bearer.

In the next step, the P-GW uses the TEID to identify the proper tunnel to forward the packets to the correct S-GW (the one that serves the destination UE). For instance, in Figure 3.10, considering the email application, the TEID is equal to 17.

Similarly, when the packets reach the S-GW, it looks up the destination eNodeB and the next TEID (5 for the example) to forward the packets. An analogous process happens in reverse, on the uplink.

3.10 Summary

This chapter provided an overview of how the LTE base station shares its resources among the mobiles. We started presenting the physical layer aspects, highlighting that OFDMA is an important tool to avoid the ISI and to facilitate the use of advanced techniques of resource allocation. We also showed that despite the advantages of OFDMA, it is not suitable for the uplink, because of the large variation in the power of the

transmitted signal. In this sense, we presented the SC-FDMA as a more appropriated technique for the LTE uplink and we highlighted the main differences of these two access methods.

Posteriorly, we presented the LTE resource grid, which is the same for downlink and uplink. The resource grid is formed by the junction of two domains: frequency and time. In this sense, an important entity of the resource grid is the Resource Block, which is composed by 12 subcarriers (180 kHz each) in frequency, by one slot (0.5 ms) in time.

Another important aspect of LTE presented in this chapter is the Link Adaptation, which is a mechanism used to adjust the modulation and coding rate to meet the current channel capacity.

We also showed how the LTE network recover from losses using HARQ schemes and how the UE notifies the eNodeB that it has data to transmit using the BSR.

Finally, we ended this chapter studying the QoS mechanisms, presenting the EPS bearer structure and the different QoS Class Identifiers.

In the next chapter, we will see how these concepts are applied in the uplink packet scheduling and which are the most relevant uplink scheduling algorithms present in the literature.

"An expert is a person who has made all the mistakes that can be made in a very narrow field."

Niels Bohr

4

LTE Uplink Packet Scheduling

Scheduling Algorithms are the entities responsible for performing the distribution of the network resources among the users. In this chapter, we present the main features of an LTE uplink scheduling algorithm. We start by stating the considerations and challenges that should be taken into account when allocating resources to UEs. Then, we introduce some of the most relevant uplink scheduling algorithms present in the literature. In Chapter 6, these algorithms will be considered as references to assess the performance of the algorithm proposed in this research.

4.1 Packet Scheduler

The Packet Scheduler (PS) is an entity of the MAC layer situated in the eNodeB. This entity is responsible for assigning the RBs to UEs, according to a particular algorithm. The 3GPP LTE standard does not specifies this algorithm, to promote different proposals among the vendors. Therefore, there are several possible solutions for the resource scheduling problem. However, generally, there are some factors that are used to design an uplink scheduling algorithm. Among them, we can cite:

- **Buffer Status Report (BSR):** each UE reports the status of its buffer to the

eNodeB. Although the report does not reflect the actual value of the buffer in bytes, but an approximate value, the eNodeB can use this information to infer for how long a packet is waiting in the buffer.

- **Bearer QoS parameters:** the scheduler may use the QoS parameters information (GBR, delay, packet loss ratio) to decide how the resource will be allocated.
- **HARQ retransmissions:** the scheduler must know which RBs are already scheduled for retransmission and which RBs are free for new transmissions.
- **Sounding measurements:** as we saw in Chapter 3, the uplink channel quality assessment is performed by the Sounding Reference Signals. The channel quality information is a very important factor to allow the scheduler to make use of the network resources in an effective manner.
- **Power Headroom Reports (PHR):** UEs present limited battery life, which affects the uplink transmission power and the scheduling process.
- **Contiguity constraint:** as explained in Section 3.1.2, SC-FDMA scheme requires that the RBs allocated to a particular UE must be contiguous[1].

Hence, the design of a scheduling algorithm can consider some or all these factors, depending on its goals. Scheduling algorithms are designed to meet the desired system objectives. Examples of these objectives are:

- **Achievable throughput:** the packet scheduler tries to use the radio interface resources as efficiently as possible, maximizing the amount of data successfully transmitted over the air interface.
- **Fairness:** the packet scheduler tries to equally share the resources among the users.
- **QoS satisfaction:** the packet scheduler tries to meet the QoS requirements of the applications, such as minimizing packet delay and packet losses.
- **Power utilization:** the packet scheduler attempts to minimize the per UE power utilization on the uplink.

In this sense, the packet scheduling can be considered as an optimization problem, in which the scheduler should find the best mapping between the RBs and UEs, in a manner that one of the objectives cited above is maximized.

For the uplink case, this mapping is even more challenging, since the RBs allocated to a particular UE must be contiguous. According to authors in [5], the contiguous RB

[1]From Release 10, the RBs allocated to a particular user can be grouped in up to two clusters of RBs.

allocation constraint is sufficient to make the problem NP-hard, i.e., it is not practical to perform an exhaustive search.

Considering the complexity of the problem, the design of a packet scheduler can be divided into two steps [34]:

- **Utility function:** it is a mathematical model used to guide the scheduling process towards meeting the target requirements. In other words, it is a *metric* to measure how close or far the scheduling process is from the desired objectives. The metric value can be distinct for each pair UE/RB, at each instant. This allows the scheduler to evaluate the several possible different mapping solutions.
- **Search algorithm:** it is the tool used to search for the solution that best optimizes the scheduler's utility function.

Once the utility function and the search algorithm scheme were chosen, the next step is the implementation. Table 4.1 shows a generic metric matrix. This matrix presents U users and M RBs. The variable λ indicates the values of the metric for each user in each RB. The calculation of λ will depend on the chosen utility function. Then, this matrix is passed to the search algorithm, which will look for the best combination of UEs/RBs, based on the metric values.

Sometimes, the process is decoupled in two stages: a Time Domain Packet Scheduling (TDPS), and a Frequency Domain Packet Scheduling (FDPS). This strategy can be interesting to minimize the complexity of the search algorithm. In this scenario, the TDPS performs a prioritization of the UEs, i.e., a shorter list of UEs with highest metric values. Hence, this reduced list is passed to the FDPS, which is responsible to perform the UE-to-RB mapping.

Table 4.1: Metric matrix.

	RB_1	RB_2	$\cdots$	RB_M
UE_1	$\lambda_{1,1}$	$\lambda_{1,2}$	$\cdots$	$\lambda_{1,M}$
UE_2	$\lambda_{2,1}$	$\lambda_{2,2}$	$\cdots$	$\lambda_{2,M}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
UE_U	$\lambda_{U,1}$	$\lambda_{U,2}$	$\cdots$	$\lambda_{U,M}$

4.2 Utility Functions

As said in the last section, utility functions are mathematical models related to the target requirements. These target requirements can be expressed in a set of parameters, such as instantaneous data rate, past average throughput, head of line packet delay, target

delay, target packet loss ratio, and more. These parameters are widely used for the definition of metrics. The ones used in this thesis are detailed in Table 4.2.

Table 4.2: Notation used for scheduling metrics.

$\lambda_{u,m}(t)$	Generic metric of the u-th user on the m-th RB at time t.
$r_u^m(t)$	Achievable data-rate for the u-th user on the m-th RB at time t.
$\bar{R}_u(t)$	Past average throughput achieved by the u-th user until time t.

In the literature, one can find several different utility functions strategies. In this research, we are particularly interested in the two most classical utility functions: Maximum Throughput and Proportional Fairness metrics. This choice was made since the principles of these two utility functions are the base of many other utility functions. Interested readers can refer to [1] for more details about other utility function strategies.

4.2.1 Maximum Throughput (MT)

The Maximum Throughput strategy aims to maximize the overall throughput by assigning the RB to the user that can achieve the maximum throughput in the current TTI. Mathematically, this metric can be expressed as:

$$\lambda_{u,m}^{MT}(t) = r_u^m(t) \tag{4.1}$$

where $r_u^m(t)$ can be calculated using the AMC module or estimated by the channel capacity.

Despite the good performance for the aggregated cell throughput, the MT scheme offers an unfair resource sharing, since UEs with poor channel quality can suffer of starvation issues.

4.2.2 Proportional Fairness (PF)

The Proportional Fairness strategy is commonly employed to find a good trade-off between fairness and spectral efficiency. The PF metric can be expressed as:

$$\lambda_{u,m}^{PF}(t) = \frac{r_u^m(t)}{\bar{R}_u(t-1)} \tag{4.2}$$

Then, using the past average throughput, users in bad conditions will be surely served within a certain amount of time.

In the next section, we present some of the most relevant uplink scheduling algorithms found in the literature. The algorithms will be briefly described and we also present a pseudocode to facilitate the understanding. The notations and nomenclatures used in the pseudocodes are detailed in Table 4.3. These algorithms were chosen based on their unique design, which has been used as base of several studies in the LTE uplink resource allocation.

Table 4.3: Pseudocode notation and nomenclature.

Symbol	**Description**
U	the number of UEs that are available for scheduling at a given TTI.
$\mathcal{U}$	the set of all UEs that are available for scheduling at a given TTI, where $\mathcal{U} = \{1, 2, ..., u, ..., U\}$.
M	the number of RBs that are available for scheduling at a given TTI.
$\mathcal{M}$	the set of all RBs that are available for scheduling at a given TTI, where $\mathcal{M} = \{1, 2, ..., m, ..., M\}$.
Λ	the metric matrix with dimension UxM.
$\Lambda(u, m)$	the scheduling metric for UE u on RB m. It is the same as $\lambda_{u,m}$.
$\Lambda(u, :)$	the scheduling metrics of all RBs for a particular UE u.
$\Lambda(:, m)$	the scheduling metrics of all UEs at a particular RB m.
$\mathcal{I}_u$	set of RBs already assigned to user u.
m_u^l	most left RB assigned to user u.
m_u^r	most right RB assigned to user u.
$\lvert \cdot \rvert$	the vertical line notation indicates the number of elements of a given set.
$\lfloor \cdot \rfloor$	floor function.
$a \leftarrow b$	Assign b to a.
$\mathcal{S} - \{s\}$	remove element s from the set $\mathcal{S}$.

4.3 Round Robin (RR)

Round Robin (RR) is a classical scheduling algorithm. Its main features are the simple design, easy implementation and immunity to starvation. In the classical approach, the RR scheduling algorithm divides the available RBs in equal portions and assigns them in circular order, handling all UEs without priority. Therefore, it is a channel-unaware

scheduling algorithm, i.e., it does not consider the channel state to assign the Resource Blocks. In this sense, the RR scheduler generally presents a lower throughput, when compared to channel-aware schedulers.

Algorithm 1 Round Robin (RR)

1: Divide available RBs into groups of Resource Chunks (RC) according to $\left\lfloor \frac{|\mathcal{M}|}{|\mathcal{U}|} \right\rfloor$
2: Distribute RCs among available UEs in an even fashion

From Algorithm 1, one can see that sometimes the RR algorithm waste some RBs, when the division $\frac{|\mathcal{M}|}{|\mathcal{U}|}$ produces a remainder.

4.4 Riding Peaks (RP)

The Riding Peaks (RP) algorithm was proposed by Suk-Bok Lee *et al.* and it is described in [5]. This algorithm is based on the fact that there is a correlation of the channel SNR values in both time and frequency domains. This means that if a user u has a good channel quality on the RB m, there is a high probability that it has also a good channel quality on the neighboring RBs, i.e., $m-1$ and $m+1$.

In this sense, the idea of this algorithm is to "ride user's peaks" in frequency domain, by exploiting such correlations [5], as depicted in Figure 4.1.

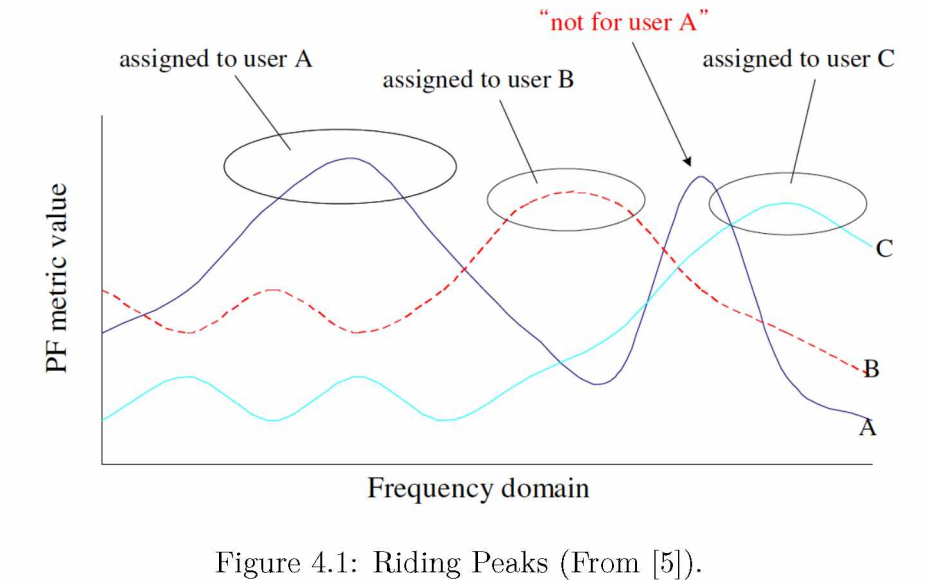


Figure 4.1: Riding Peaks (From [5]).

From Algorithm 2, one can note that Riding Peaks is a channel-aware algorithm, since it is based on the PF utility function. This algorithm is simple, but provides good results. On the other hand, it does not offer the necessary conditions to support real time applications with more strict requirements.

Algorithm 2 Riding Peaks (RP) [5]

Let $\mathcal{V}$ be the sorted list of all metrics values $\lambda_{u,m}$ in decreasing order, where $\lambda_{u,m}$ is calculated from Equation 4.2
Let $\mathcal{M}$ be the set of not-yet-assigned RBs
$k \leftarrow 1$
while $\mathcal{M} \neq \varnothing$ **do**
    Pick RB m with the k-th largest metric value $\lambda_{u,m} \in \mathcal{V}$, $m \in \mathcal{M}$
    Let $\mathcal{I}_u$ be the set of RBs already assigned to user u
    if (m is adjacent to $\mathcal{I}_u$) **or** ($\mathcal{I}_u = \varnothing$) **then**
        $\mathcal{I}_u \leftarrow \mathcal{I}_u \cup \{m\}$ ▷ Assign RB m to user u.
        $\mathcal{M} \leftarrow \mathcal{M} - \{m\}$ ▷ remove RB m from the set of available RBs.
        $\mathcal{V} \leftarrow \mathcal{V} - \{\lambda_{u,m}\}$ ▷ remove metric $\lambda_{u,m}$ from the set of the sorted metrics.
        $k \leftarrow 1$
    else
        $k \leftarrow k + 1$
    end if
end while

4.5 Recursive Maximum Expansion (RME)

The Recursive Maximum Expansion (RME) algorithm was proposed in [11] by Temiño *et al.* RME is also a channel-aware algorithm based on the Proportional Fairness metric. This algorithm is an improvement of the First Maximum Expansion (FME) algorithm proposed by the same authors.

In each iteration, the algorithm searches for the UE/RB pair with the maximum metric value, as showed in Figure 4.2. Then, it expands the RB allocation on both sides, in frequency domain, till it finds another UE with a better metric in that RB. This served UE is put into an idle mode, while the remaining RBs are distributed among the other UEs. In the case that all UEs are served, but there are some unscheduled RBs, the algorithm starts the recursive step. In this phase, the remaining RBs are assigned to the idle users in a manner that the contiguity criteria is maintained. Algorithm 3 provides the details of this scheduling algorithm.

Since RME is a PF-based scheduler, it does not provide mechanisms to ensure the QoS requirements of real time applications.

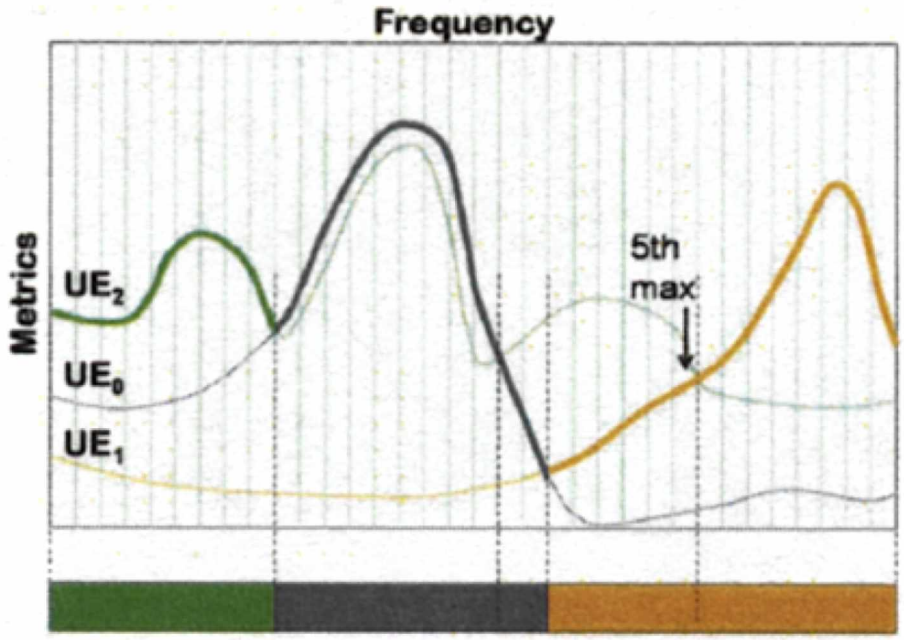


Figure 4.2: Recursive Maximum Expansion (From [11]).

4.6 Riding Peaks with QoS (RPQoS)

Despite their relevance and good results, RP and RME were designed and evaluated in simple scenarios, i.e., scenarios with only one type of traffic, which does not have any restrictions of maximum delay or minimum throughput, for example.

In this sense, to provide a fair comparison with the algorithm proposed in this thesis, it was necessary to choose a channel-aware/Qos-aware and more competitive scheduling algorithm to participate in the performance evaluation described in Chapter 6.

In [15], Safa *et al.* present a scheduling algorithm with a channel-aware/QoS-aware utility function and a searching algorithm based on RP. Since this algorithm uses the RP searching strategy combined with a QoS-aware utility function, it will be referenced as RPQoS.

The RPQoS utility function is based on the PF metric, since the authors planned to provide QoS and fairness in the resource allocation.

As explained in Section 3.7, in LTE, data are classified into four groups called Radio Bearer Groups (RBGs), each of which has a different priority level, according to Table 3.2. RPQoS makes use of this information to perform the resource allocation according to the priorities of the RBGs.

Algorithm 3 Recursive Maximum Expansion (RME) [11]

Let $\mathcal{S}$ be the list of users that have already been served by the scheduler
$\mathcal{S} \leftarrow \varnothing$
for all $u \in \mathcal{U}$ **do**
    Calculate metric $\Lambda(u, m)$, $\forall m \in \mathcal{M}$, according to Equation 4.2
end for
while $(\mathcal{M} \neq \varnothing)$ **and** $(\mathcal{U} \neq \varnothing)$ **do**
    Find UE $u \in \mathcal{U}$ and RB $m \in \mathcal{M}$ with maximum metric in Λ
    $\mathcal{I}_u \leftarrow \mathcal{I}_u \cup \{m\}$ ▷ Assign RB m to user u.
    $\mathcal{M} \leftarrow \mathcal{M} - \{m\}$ ▷ remove RB m from the set of available RBs.
    $\Lambda(u, m) \leftarrow 0$
    Let $m_l \leftarrow (m - 1)$ ▷ RB in the left.
    Let $m_r \leftarrow (m + 1)$ ▷ RB in the right.
    while $(\Lambda(u, m_l) = \max(\Lambda(:, m_l)))$ **and** $(m_l \geq 1)$ **do** ▷ Going left.
        $\mathcal{I}_u \leftarrow \mathcal{I}_u \cup \{m_l\}$
        $\mathcal{M} \leftarrow \mathcal{M} - \{m\}$
        $m_l \leftarrow (m_l - 1)$
    end while
    while $(\Lambda(u, m_r) = \max(\Lambda(:, m_r)))$ **and** $(m_r \leq M)$ **do** ▷ Going right.
        $\mathcal{I}_u \leftarrow \mathcal{I}_u \cup \{m_r\}$
        $\mathcal{M} \leftarrow \mathcal{M} - \{m\}$
        $m_r \leftarrow (m_r + 1)$
    end while
    $\mathcal{U} - \{u\}$
    $\mathcal{S} \cup \{u\}$ ▷ Adding UE to idle mode.
end while
while $\mathcal{M} \neq \varnothing$ **do** ▷ Check and distribute remaining RBs among idle UEs.
    Find UE $u \in \mathcal{S}$ and RB $m \in \mathcal{M}$ with maximum metric in Λ
    if $m = m_u^l - 1$ **then**
        Expand allocation to the left, until finding a UE with a higher metric, or a RB already allocated
    else if $m = m_u^r + 1$ **then**
        Expand allocation to the right, until finding a UE with a higher metric, or a RB already allocated
    end if
end while

In this context, the utility function of the RPQoS algorithm is defined as [15]:

$$\lambda_{u,m}^{RPQoS}(t) = \frac{\lambda_{u,m}^{PF}(t)}{\alpha_u(t)} \tag{4.3}$$

where $\alpha_u(t)$ is the QoS introducer of user u at time t. To understand how $\alpha_u(t)$ is calculated, first it is necessary to define some concepts:

- D_k: let it be 90% of the minimum of the maximum allowed delay of all applications belonging to RBG k; D_k = *min (max delay of application A, max delay of application B, ..., max delay of application Z)*, such that applications A through Z are all applications that belong to RBG k.
- $d_u^k(t)$: let it be the minimum of D_k and the time from when a burst of packets arrived to the RBG k of user u till time t; i.e., $d_u^k = min(D_k, t - startTime(burst_u))$.
- $\tau_u(t)$: let it be the delay weight; $\tau_u(t) = max(\frac{d_u^k(t)}{D_k})$ for $1 \leq k \leq 4$ and for all packet bursts of user u. It starts at 0, since $d_u^k(t)$ will be 0 at first TTI, and as data of user u approaches 90% of their maximum allowed delay, it gets closer to its upper limit 1, since $d_u^k(t)$ will be equal to D_k, when that value is reached.
- $\delta_u^k(t)$: let it be the data pending for transmission of user u at time t in the RBG k.

Then for every $\delta_u^k(t) > 0$ and for every user u, the QoS introducer, $\alpha_u(t)$ is given as:

$$\alpha_u(t) = min(k) - min(k) * \tau_u(t) + log(min(k)) + \epsilon. \tag{4.4}$$

The Equation 4.4 is divided into four parts that can be described as following:

- $min(k)$: which allows $\alpha_u(t)$ to be smaller for users having higher priority data. For example, if $\delta_u^2(t)$ and $\delta_u^3(t)$ are not 0 (i.e., UE u has data pending for transmission in RBG_2 and RBG_3 at time t), $min(k)$ would be 2.
- $-min(k) * \tau_u(t)$: which starts at 0 and reaches $-min(k)$, when data is waiting to be transmitted for 90% of their maximum allowed delay.
- $log(min(k))$: which allows the differentiation among users having different priorities if 90% of their maximum allowed delay is reached at the same time, thus giving an advantage to the highest priority users.
- ϵ which is used to never allow α to be 0.

Algorithm 4 provides details about the RPQoS approach.

From Algorithm 4, RPQoS calculates the metric matrix. Then, this matrix is the input of the RP algorithm, already described in Algorithm 2.

Algorithm 4 Riding Peaks with QoS (RPQoS) [15]

Let r_u^m be the channel quality indicator of UE u on RB m
Let R_u be the long term service rate of UE u
Let B_u^k be the buffer status of UE u in RBG k
for $u = 1$ to U **do**
    Update R_u
    $max = 0$
    $min = 0$
    for RBG $k = 4$ to 1 **do**
        Update B_u^k
        if ($\delta_u^k > 0$ **then**
            d_u^k++
            $min = k$
            if $max < \frac{d_u^k}{D_k}$ **then**
                $max = \frac{d_u^k}{D_k}$
            end if
        end if
    end for
    $\tau_u = max$
    $\alpha_u = min - min * \tau_u + log(min) + \epsilon$
    for RB $m = 1$ to M **do**
        $\lambda_{u,m}^{PF} = \frac{r_u^m}{R_u}$
        $\lambda_{u,m}^{RPQoS} = \frac{\lambda_{u,m}^{PF}}{\alpha_u}$
    end for
end for

4.7 Summary

This chapter offers an overview of the LTE uplink packet scheduling design process. We started by showing that the PS is at the MAC layer and we also presented a list of issues that should be considered when designing a scheduling algorithm.

Since the PS is a complex and important part of the network, it is generally divided into two portions: the utility function and the search algorithm. The utility function is the way used to make the scheduler meet the desired system objectives, like throughput, fairness, QoS satisfaction, etc. Then, we introduced two examples of utility functions, the MT and the PF metrics.

The next step is the definition of the search algorithm. In this sense, we presented four searching strategies: Round Robin, Riding Peaks, Recursive Maximum Expansion

and Riding Peaks with QoS. The RR algorithm is a classical algorithm chosen as a baseline to evaluate the other schedulers and represents the class of algorithms that are channel-unaware/QoS-unaware.

RP and RME algorithms were chosen since they were the most relevant algorithms found in the literature. These algorithms were cited, studied, compared and used as based of several other algorithms proposals. These two algorithm represent the class of algorithms that are channel-aware/QoS-unaware.

Despite their relevance and good results, RP and RME do not provide QoS support. In this context, RPQoS was chosen as the main competitor for our proposed algorithm and represents the class of algorithms that are channel-aware/QoS-aware.

In the next chapter, we describe a new scheduling algorithm based on Genetic Algorithms and designed to work in mixed traffic environment.

"They did not know it was impossible so they did it."

- Mark Twain

5

A New Three-Step GA-Based Scheduling Algorithm for the LTE Uplink

THE MULTIPLE ACCESS STRATEGY chosen for the LTE uplink is the SC-FDMA scheme. This choice increases the complexity involved in the resource allocation, since SC-FDMA requires that the Resource Blocks assigned to a particular user must be contiguous. In this chapter, we propose a new strategy to perform the resource allocation in the uplink. This approach is based on Genetic Algorithms (GA), which is a powerful tool designed to deal with complex optimization problems. In this sense, we start introducing the main concepts related to GA. Then, we describe the details of the proposed scheduling algorithm.

5.1 Genetic Algorithms

Evolutionary Algorithms are strategies that use computational models of the natural evolution processes as a tool for solving problems [35]. In the 1950s and the 1960s several computer scientists independently studied evolutionary systems with the idea that evolution could be used as an optimization tool for engineering problems. The

idea in all these systems was to evolve a population of candidate solutions to a given problem, using operations inspired by natural genetic variation and natural selection [18].

On this context, Genetic Algorithms were invented by John Holland in the 1970s and can be considered as a branch of the Evolutionary Algorithms. Holland presented the Genetic Algorithms as an abstraction of biological evolution and gave a theoretical framework for adaptation under the GA. This framework introduced the concept of *population*, which is formed by *individuals*. Each individual is a possible solution for the problem and it is represented by a *chromosome*, which, in turn, is formed by *genes*.

These chromosomes are submitted to genetic operations. These operations mimic phenomena seen in nature as *crossover* and *mutation*. From the chromosome, the algorithm can evaluate how good that solution is, using a *fitness function*. Thus, the algorithm can infer which solutions are the fittest, and then, decide which individuals will reproduce and pass their genetic code to the next generation. Therefore, GA is guided by the "survival-of-the-fittest" principle.

GA is considered a population based meta-heuristic. This means that it does not provide the optimal result every time the algorithm runs. In fact, it provides a good solution, but we have no assurance to obtain the best or optimal solution for that problem. This is a consequence of the design of the GA that presents operations based on randomness. However, this does not mean that GA is a blind searching strategy. It does have random components, but it employs the information of the current state to guide the search.

5.2 Three-Step GA-Based Scheduling Algorithm

GA is capable of delivering near-optimal performance with comparatively low complexity when compared with an exhaustive search. However, when compared with "greedy" scheduling algorithms, GA can present a considerable greater complexity. This factor is important, since in LTE networks, the eNB must allocate the resources each TTI, which lasts 1 ms. Then, it was necessary to create strategies to use the searching power of GA, while controlling its complexity.

In this context, we propose a Three-Step GA-based scheduling algorithm, from now on called TSGA. From preliminary tests, we concluded that GA is a promising tool for solving the resource allocation in the uplink channel. However, employing GA from a unique block approach, lead the algorithm to a slow convergence, since it invest time evaluating solutions, where the users could postpone the transmission, without reaching

the delay budget.

In this sense, we propose a scheduling process divided in multiple steps, which allows us a better control of the complexity of GA search and facilitates the provision of QoS as well. Figure 5.1 shows an overview of TSGA.

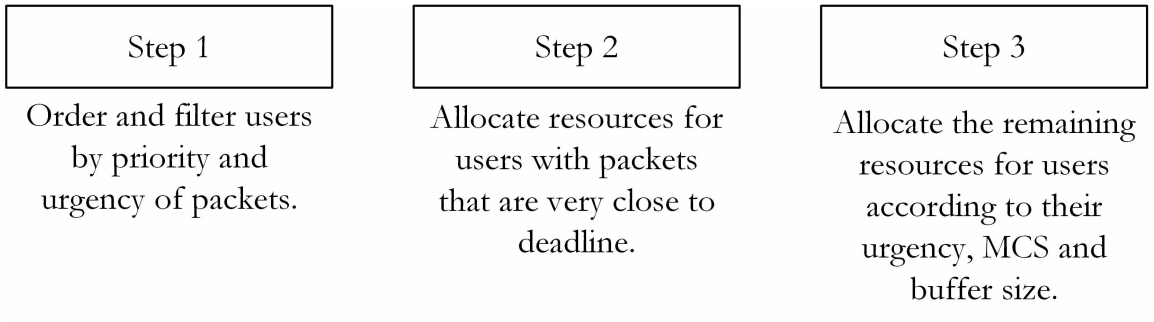


Figure 5.1: TSGA overview.

5.2.1 Step One

Step One is responsible for ordering and filtering users by priority and urgency of packets. The urgency factor is calculated from the packet delay budget defined for each QCI, described in Table 3.2. Then, we can define:

$$u_i = \frac{h}{D} \tag{5.1}$$

where u_i is the urgency of the Head of Line (HoL) packet in the buffer of user i, h is the time interval that the HoL packet is waiting in the buffer and D is the packet delay budget for the application that this packet belongs to.

To provide QoS mechanisms, we also define the priority factor of user i as:

$$p_i = \begin{cases} 2 & \text{if application is GBR} \\ 1 & \text{if application is Non-GBR} \end{cases} \tag{5.2}$$

From Equation 5.2, the algorithm gives more priority for applications with Guaranteed Bit Rate (GBR), according to QCI definition. In order to differentiate the nine priorities defined for each QCI, the algorithm employs a fine tunning from the packet delay budget (D), calculated from the urgency factor u_i. From Equations 5.1 and 5.2,

it is possible to define the metric that controls the ordering of users in Step One as:

$$\alpha_i = p_i \cdot u_i \tag{5.3}$$

The N active users are ordered according to the value of α_i. The α_i factor indicates the priority of users, including the priority differences among users that belong to different GBR classes. Then, this ordered list $\mathcal{L}$ is passed to the Second Step of the algorithm.

5.2.2 Step Two

Step Two is responsible for two tasks:

- Allocation of resources for users with packets presenting latency very close to the packet delay budget defined for the application;
- Selection of which users of the ordered list from Step One will pass to the Step Three of the algorithm.

Step Two of the algorithm evaluates each user in the ordered list $\mathcal{L}$ according to the strategy described in Algorithm 5.

From Algorithm 5, one can see that Step Two allocates resources for users with very urgent packets. The urgency threshold was set to 90% of the packet delay budget, as recommended in [15]. This strategy tries to assure the QoS requirements of the applications. The urgency factor clearly indicates to the algorithm which users need resources. In this context, it is not efficient to trigger the GA search to accommodate the resources among the users.

When user i presents $u_i \geq 0.9$, Step Two verifies all RBs, identifying which are available. A particular RB can not be available, because it was already allocated by the HARQ entity, or because it presents CQI=0 for that user i. Then, after verifying all RBs, Step Two can determine the largest interval of contiguous RBs. Hence, based on the BSR and MCS, Step Two calculates the demand of RBs for user i and allocates those contiguous RBs according to the demand d_i.

On the other hand, when the urgency factor is below 0.9, the user is selected for Step Three and will dispute the resources with other users. Here, Step Two tries to control the complexity of the GA search, limiting the number of users in Step Three to 5. Step Two is also responsible for avoiding starvation in the scheduling, since it allocates resources for the very urgent users.

Algorithm 5 Step Two

Let $\mathcal{L}$ be the ordered list of users from Step One.
Let $\mathcal{V}$ be the set of users selected for Step Three.
Let $\mathcal{M}$ be the set of RBs.
Let $\mathcal{C}$ be a set of contiguous RBs in $\mathcal{M}$.
Let $\mathcal{I}$ be a set of vectors $\mathcal{C}$.
Let d_i be the demand of RBs of user i.
while $\mathcal{L} \neq \varnothing$ **or** $\mathcal{M} \neq \varnothing$ **or** $|\mathcal{V}| < 5$ **do**
    Pick the first user i in $\mathcal{L}$.
    if $u_i \geq 0.9$ **then**
        for all $k \in \mathcal{M}$ **do**
            if k is available **then**
                $\mathcal{C} \leftarrow \mathcal{C} + \{k\}$ ▷ Append RB k.
            else
                $\mathcal{I} \leftarrow \mathcal{I} + \{\mathcal{C}\}$ ▷ Append set $\mathcal{C}$.
                $\mathcal{C} \leftarrow \{\}$ ▷ Empty set $\mathcal{C}$.
            end if
        end for
        $\mathcal{C}_{largest} = max(\mathcal{I})$ ▷ Find the largest set $\mathcal{C}$ in $\mathcal{I}$.
        if $d_i < |\mathcal{C}_{largest}|$ **then**
            Allocate d_i RBs in $\mathcal{C}_{largest}$ to user i.
            Remove d_i RBs in $\mathcal{C}_{largest}$ from set $\mathcal{M}$.
        else
            Allocate all RBs in $\mathcal{C}_{largest}$ to user i.
            Remove all RBs in $\mathcal{C}_{largest}$ from set $\mathcal{M}$.
        end if
    else
        Add user i to set $\mathcal{V}$.
    end if
    Remove user i from set $\mathcal{L}$
end while

The values of urgency threshold and number of users in Step Three were defined in an empirical manner. For urgency threshold, if the value is too close to 1, maybe the time interval is too short to transmit the packet in the buffer in time. On the other hand, if the value is close to 0, Step Two prematurely allocates the majority of RBs and the algorithm will lose the power and advantages of the GA search. In what concerns the number of users selected for Step Three, if this value is high, the GA search will be more complex since the searching space will be increased. If this value is too small, the users may not use all the RBs and the allocation would not be efficient. Ordering and

limiting the number of users decreases the complexity of the GA search, while indicating a more accurate searching space.

5.2.3 Step Three

Step Three is responsible for allocating the remaining resources among the users selected in Step Two. As said before, this allocation is based on Genetic Algorithms. As a population-based metaheuristic, GA makes use of the concept of individuals. The individuals of the population are submitted to a set of genetic operations: representation, initialization, evaluation, selection, recombination and termination. Then, a new population is derived from the older one. The principle of survival of the fittest ensures the evolution of the solutions as the algorithm progresses. Figure 5.2 shows how GA works.

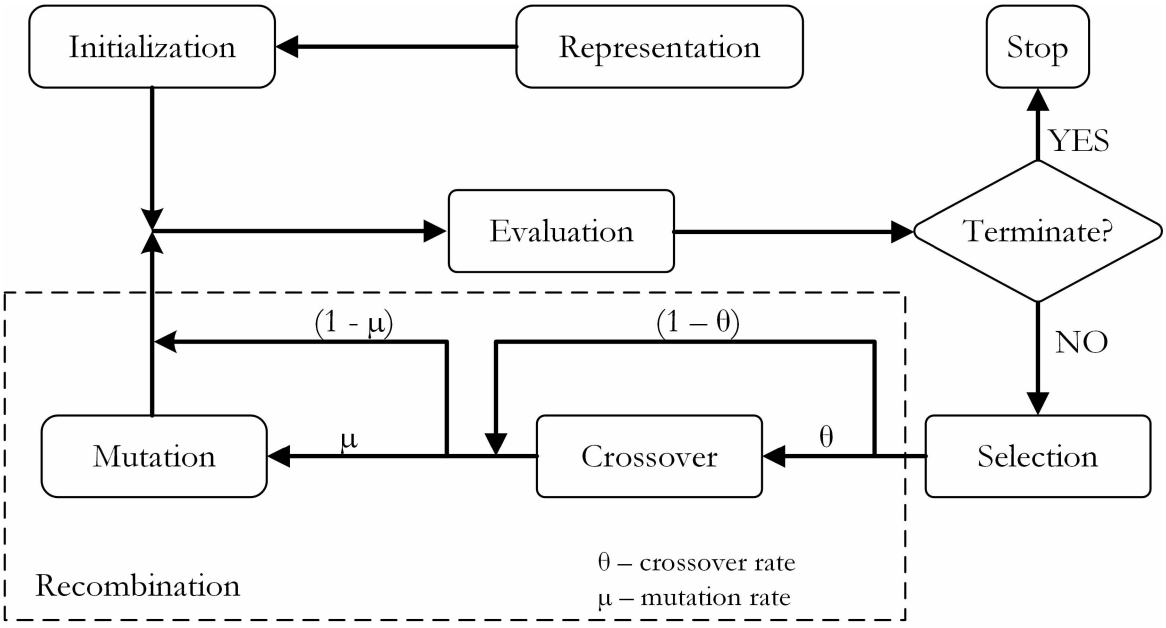


Figure 5.2: GA flowchart: how it works.

Next, each of the genetic operations indicated in Figure 5.2 will be described.

Representation

In order to use Genetic Algorithms to find a solution for the resource allocation problem, the first step is to code the solutions into sequences of genes, also known as chromosome. This step is called *representation*.

In the literature, one can find some strategies to perform this mapping. For example,

in [3], the authors use a binary matrix to represent the RB assignment. In this scheme, we have a matrix $\mathcal{V}$ of dimensions U x M, i.e., U users by M RBs. Then, whenever a RB m is assigned to a particular UE u, the position $v_{u,m}$ of matrix $\mathcal{V}$ receives 1 as value. Figure 5.3 illustrates the mapping process.

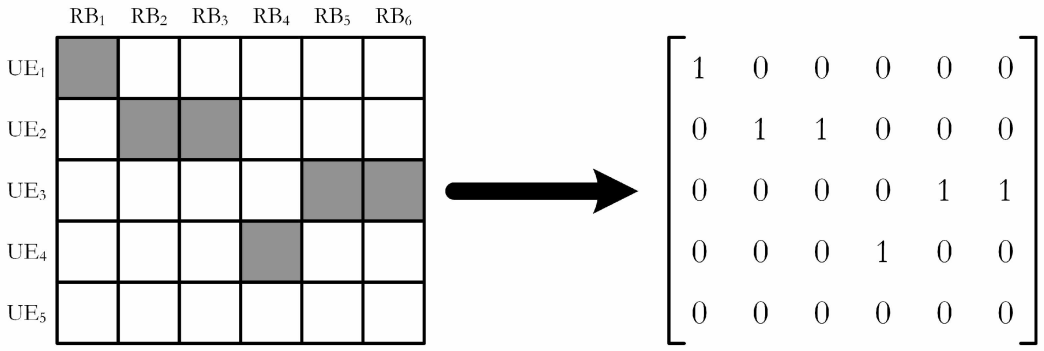


Figure 5.3: Binary matrix representation.

The binary matrix mapping presents a drawback. Since it is based on a matrix, the implementation may present issues concerning the scalability. The greater the number of users and RBs, the greater the time to process the matrix. Generally, matrices are implemented with a chain of *for* statements, which are time-consuming operations. In this sense, it is necessary to simplify the binary matrix approach. This is done in [19], where the authors transformed the binary matrix into a vector. Instead of assigning 1 and 0 values, they used labels representing each user to form a vector.

The complexity of the vector approach still depends on the number of RBs, but it does not depend on the number of users. This fact reduces the complexity of the algorithm, when compared to the binary matrix approach.

In this sense, the simplest way to code the resource allocation map is considering each resource block as a gene. Each gene can take the value of a particular user, indicating that the RB has been granted to that user. Then, we defined a chromosome based on the vector representation to use in our algorithm. Figure 5.4 illustrates a chromosome example, considering 6 RBs and 5 users. This figure also highlights the *boundaries* between the RBs assigned to different users. This information will be important in the next genetic operations.

From Figure 5.4, one can note that the chromosome is derived from the matrix formed by the available users and RBs. First, all users and RBs are available for the scheduling process. However, after a while, the system may start to experience losses. This will trigger the HARQ entity to recover from these losses. In this scenario, the HARQ entity will reserve some RBs to perform the retransmission. First, it will try to retransmit in the same RBs of the first transmission. If it is not possible, HARQ entity

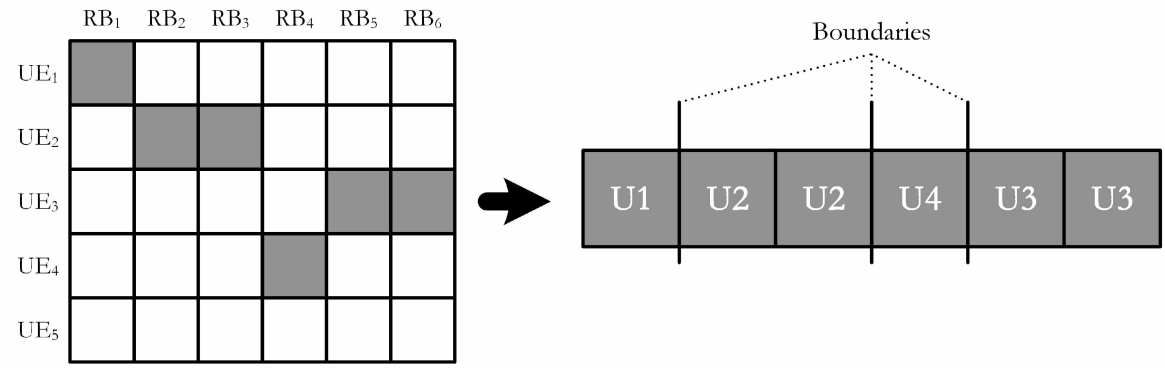


Figure 5.4: Vector representation.

will search for new suitable RBs.

The scheduling process is performed after the HARQ allocation. Then, sometimes, not all RBs will be available for scheduling of new transmissions. Another important detail to advert to: when a user is scheduled by HARQ entity for retransmission, the scheduler does not consider that user available for transmission of new data in that TTI. This is not a constraint of the LTE standard, but only a convention or simplification adopted by the simulation environment. There is also another case, when a particular UE has a CQI equal to zero to any of the RBs. This means that the user is out of the range of the base station and it should be excluded from the scheduling process. A third case is also possible, when we have UEs using the PRACH to get access to the eNB. This is a typical case of the beginning of the simulation, when some RBs are allocated to the establishment of the connection of new users in the cell. All these scenarios cited above impact on the composition of the metric matrix and, consequently, in the chromosome definition.

To be clearer, let's consider the example depicted in Figure 5.5. In a given moment, we have a grid with five users and six RBs. However, the fourth RB is already scheduled to UE_3 for HARQ retransmission. One can also see that UE_5 presents $CQI = 0$ for RB_6, meaning that this user is out of range. Then, it is possible to build a new grid, considering only the available users and RBs. To facilitate the subsequent genetic operations, it is interesting to perform a mapping of the grid with only the available resources to a new grid with a sequential numbering, as depicted in Figure 5.5. This mapping must also consider the *discontinuity* produced by the HARQ retransmission, which creates *intervals of scheduling* in the grid. In the example, there are two intervals and the scheduler must know that it must not assign RB_3 and RB_4 of the new grid to the same user, since these RBs are not contiguous.

In the next section, we delineate how this chromosome derived from the new grid must be initialized.

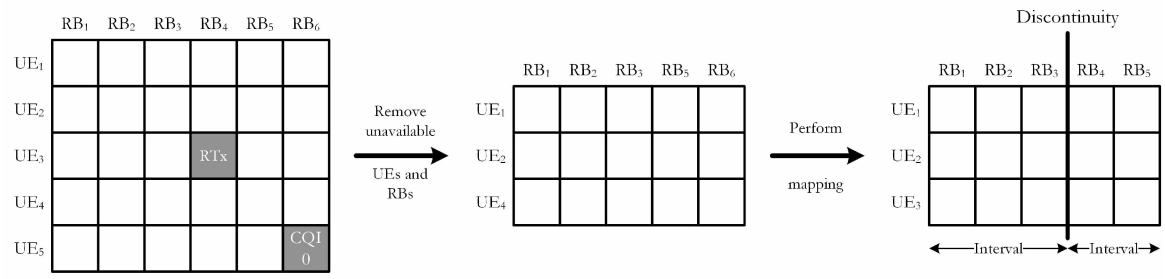


Figure 5.5: HARQ impact on the metric matrix.

Initialization

After defining the chromosome format to map the solutions, the next step is the *initialization* of the chromosome. In the downlink, it is common to assign a random user to each RB/gene of the chromosome. However, this approach is not suitable for the uplink, since it does not consider the contiguity constraint and, so, it produces a significant number of infeasible solutions. In this sense, we had to elaborate an initialization scheme to our GA-based algorithm. Algorithm 6 describes this scheme.

As we said in the last section, the retransmissions scheduled by the HARQ entity produce discontinuities in the chromosome. These discontinuities will create intervals of scheduling. This means that a given user cannot receive RBs that belong to different intervals, since this would break the contiguity. Despite the fact that from Release 10, the eNB can assign non contiguous RBs to a particular user, our initialization algorithms still will be needed, since this non contiguous assignment must not be greater than two frequency-separated groups of resource blocks. In this thesis, due to restrictions of the simulation tool, we are considering the Release 8. Hence, the initialization algorithm will assign only one contiguous interval of RBs for each user.

In this sense, the initialization algorithm starts by identifying the intervals and the size of each of them in number of RBs (line 1). We must also identify which users are available to receive RBs (line 2).

Then, we pass through each of the intervals to allocate their RBs. To perform this task, we identify the number of intervals and compare it with the number of available users (line 6). If we have more intervals than available users, then, we will select only one user for the current interval (line 7). The remaining users will be allocated in the subsequent intervals, to better explore the chromosome. Otherwise, if we have more users than intervals, we save some users for the subsequent intervals, and allow the rest of users to dispute the RBs of the current interval (line 9).

The next step is the assignment of the RBs of the current interval (line 11). If we

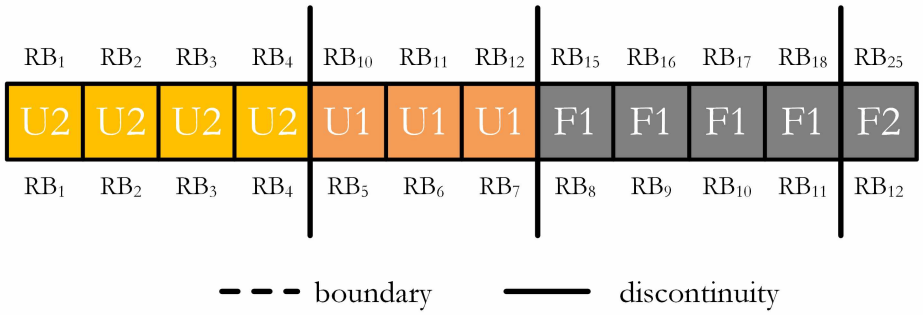
Algorithm 6 GA initialization algorithm

1: Let $\mathcal{I}$ be a vector with the size of each interval in the chromosome.
2: Let $\mathcal{U}$ be the set of available users.
3: $k \leftarrow |\mathcal{I}|$ ▷ k: number of intervals to allocate.
4: **for all** $i \in \mathcal{I}$ **do**
5: $r \leftarrow i$ ▷ r: RBs to allocate in the interval.
6: **if** $|\mathcal{U}| < k$ **then**
7: $j \leftarrow 1$ ▷ j: number of users allowed to receive RBs.
8: **else**
9: $j \leftarrow |\mathcal{U}| - (k - 1)$
10: **end if**
11: **while** $r > 0$ **do**
12: **if** $\mathcal{U} \neq \varnothing$ **then**
13: Choose a random user u in $\mathcal{U}$.
14: **if** $j = 1$ **then**
15: Only one user. Assign all RBs in the interval to user u.
16: $r \leftarrow 0$
17: **else**
18: Choose a random number c in the interval $[1, r]$.
19: Assign c RBs to user u.
20: $r \leftarrow r - c$
21: **end if**
22: $j \leftarrow j - 1$
23: Save the location of the boundary between user u and the next user that will receive RBs.
24: $\mathcal{U} \leftarrow \mathcal{U} - \{u\}$
25: **else**
26: Assign the remaining RBs to an abstract user to keep the size of the chromosome constant.
27: **end if**
28: **end while**
29: $k \leftarrow k - 1$
30: **end for**

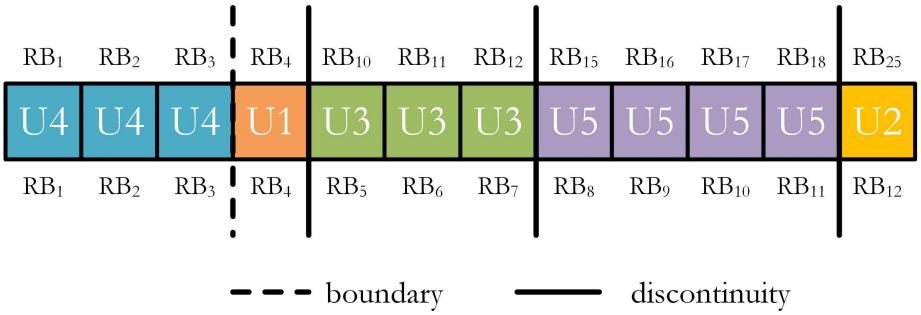
have users to assign RBs, then, we choose a random user u among the available users (line 13). Then, according to the number of allowed users, we assign the RBs. If we have only one allowed user, the user u takes all the RBs of the current interval (line 15). Otherwise, we choose a random number of RBs to assign to user u (line 18). In both cases, after the assignment process, we remove the assigned RBs from the set of available RBs.

In the next step, we decrement the number of users allowed to receive RBs and remove the current user u from the set of available users (lines 22 and 24). We also save the location of the boundary between user u and the next user that will receive RBs (line 23). This information is important for the subsequent genetic operations.

Finally, when all available users have received RBs, but we still have RBs to assign, we must assign these remaining RBs to abstract users. This means that these RBs will be assigned to *fake* users, with the purpose of *filling* the chromosome and keep its size constant and independent of the number of discontinuities created by the HARQ process (line 26). A fixed size chromosome facilitates the operations of crossover and mutation. To clarify these scenarios, Figure 5.6 shows two cases of chromosome initialization.



(a) 2 users and 4 intervals of scheduling.



(b) 5 users and 4 intervals of scheduling.

Figure 5.6: Utilization examples of the initialization algorithm.

In Figure 5.6a, we have two available users to receive RBs. Note that there are discontinuities in the resource grid. In the upper labels, we have the real RB positions. In the lower labels, we have the sequential mapping. These discontinuities created 4 intervals of scheduling. Then, in this case, we have more intervals than available users. In this context, we have to fulfill the chromosome with fake users to keep the size of chromosome constant.

On the other hand, in the case of Figure 5.6b, we have more users than intervals. Then, we do not have to use fake users and some intervals will be shared by more than one user, creating boundaries in the chromosome.

Most of time, the system will experience the case of Figure 5.6b, since there will be many users requesting resources from the network. On the other hand, the case of Figure 5.6a is typical of the beginning of the simulation. In this sense, the initialization algorithm must be prepared to deal with all type of these situations.

Evaluation

In the evaluation phase, each individual of the population is evaluated according to a fitness function. The fitness function is related to the problem we are trying to solve with GA, i.e., it is the tool used to meet the desired system objectives.

Scheduling algorithms are generally composed of two parts: the utility function and the searching algorithm. In GA-based scheduling algorithms, the fitness function is responsible for the utility function role.

Then, to define the utility function, first, we have to establish what is the system objective that we are interested. In this thesis, we are concerned with Video Chat transmission. This type of traffic was chosen since video transmission is one of the most important applications in the Internet and it presents strict requirements to offer a satisfactory service. Among the types of video transmission, we selected Video Chat, also known as Video Call. Since we are working with the uplink, Video Chat transmission requires more of the uplink resources than Video Streaming, for example, and, therefore, it presents a more challenging environment for the uplink scheduling algorithms.

In Video Chat transmission, one important factor is the packet delay experienced by the two sides of the connection. For video transmission with good quality, the packets must arrive within a certain interval. If a packet arrives after this threshold, it is not important anymore, from the receiver point of view. In this context, to get a better video transmission quality, we must keep the packet latency below the packet delay budget defined for the application. This should happen in a scenario with UEs running other types of applications as VoIP and FTP.

In video transmission, the *video sequence* is divided into a set of *frames*, which, in turn, are also divided into a set of *packets*. When a particular UE has a new packet to transmit, it notifies the eNodeB using a BSR message. Then, as soon as the eNodeB receives the notification of data in the UE's buffer, it records this instant and the size of the packet that just arrived in the buffer.

From the BSR, the eNB can estimate the demand of bytes waiting to be transfer. In the uplink, the eNB can also estimate the user channel quality using the SRS strategy. The SRS is mapped to a SINR, and consequently to CQI and MCS values. From

the MCS, the eNB can calculate how many bytes each RB can carry, using the Table 7.1.7.2.1-1 in [36]. Then, the eNB can get a measure of the demand of RBs for the users. Finally, it is possible to evaluate each chromosome using the following fitness function:

$$fitness = \frac{1}{\sum_i u_i \cdot e_i} \tag{5.4}$$

$$e = \begin{cases} |g - d| & \text{if } g < d \\ \dfrac{|g - d|}{2} & \text{if } g \geq d \end{cases} \tag{5.5}$$

where u_i is the urgency of the HoL packet of user i, e_i is the error committed by the chromosome concerning the demand of RBs d and the RBs actually granted g. The parameter g is obtained by counting the number of genes in the chromosome assigned to user i. If $g < d$, the algorithm is granting less RBs than the demand of the user. So, this solution is worse than the case when $g \geq d$ and the user is receiving more RBs than needed. The urgency factor is included to indicate to the algorithm that it should avoid errors for the most urgent users. The smaller the error of the chromosome, the greater will the value of the fitness function be. In this sense, each individual can be evaluated with the Equation 5.4.

Selection

In this step, we select the individuals that will produce the *offspring*, i.e., the individuals of the next generation. The selection should consider the potential of the individual to create good offspring. In this research, we employ the well known *Tournament* method to select the individuals that will be submitted to recombination.

In the Tournament method, the first step is to define the number of individuals that will compete in the tournament. After that, we randomly choose the participants of the tournament among the individuals of the population. Then, we start the competition and the individual with the best fitness is chosen to the next phase of the algorithm. If we have a population of N individuals, so, we repeat the Tournament process until we have selected N individuals.

At the end of each generation, we also employ another selection method called *elitism*. In our algorithm, this method selects the best individual in the current generation to pass to the next generation. This action is important to assure that the relevant features of the individuals, selected so far by the survival-of-the-fittest principle, will not be discarded during the process of recombination.

Crossover

The crossover is part of the recombination phase. Crossover mimics the mechanism of reproduction in real world. Two different parents are chosen from the group formed in the selection phase. From these parents, two offspring are derived by recombining parts of the parents chromosome. Due to the peculiarities of resource allocation in LTE uplink, it is not possible to use classical strategies of crossover which use *switching point*. These classical strategies would generate a huge number of infeasible solutions. To solve this problem, we developed a crossover strategy that recombines the parents chromosome, while considering the contiguity constraint. Figure 5.7 shows an example of this crossover operation.

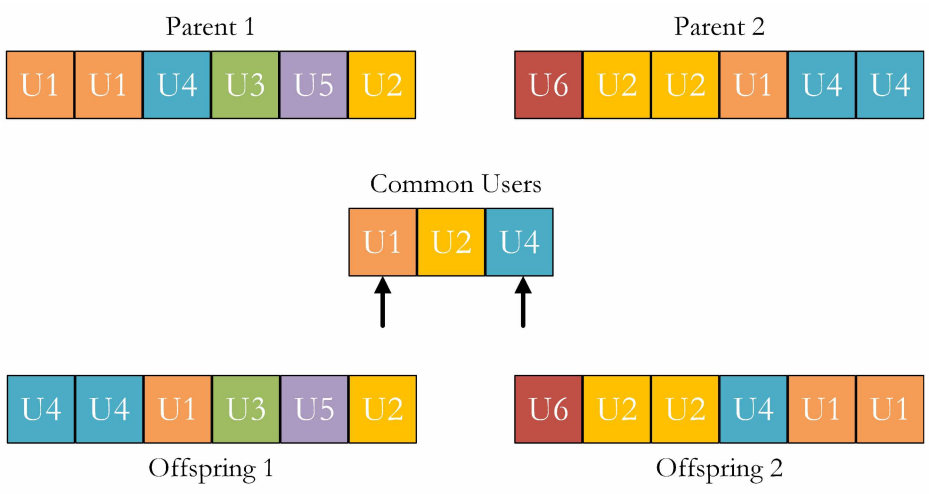


Figure 5.7: Crossover operation.

From the chosen parents, we identify which users are present in both chromosomes. If the algorithm could not find at least two common users, the crossover is not performed, and the parents pass to the next generation. In the example of Figure 5.7, we have three common users (U1, U2 and U4). From this group, we randomly select two users. In the example, we selected U1 and U4. In the next step, there is a permutation: the RBs granted to user U1 are now granted to user U4 and vice-versa. The two new chromosomes are the offsprings for the next generation.

The process is analogous in the scenario where we face discontinuities and the use of fake users. These abstract users are considered in the process of recombination of the parent's chromosomes, as depicted in the example of Figure 5.8.

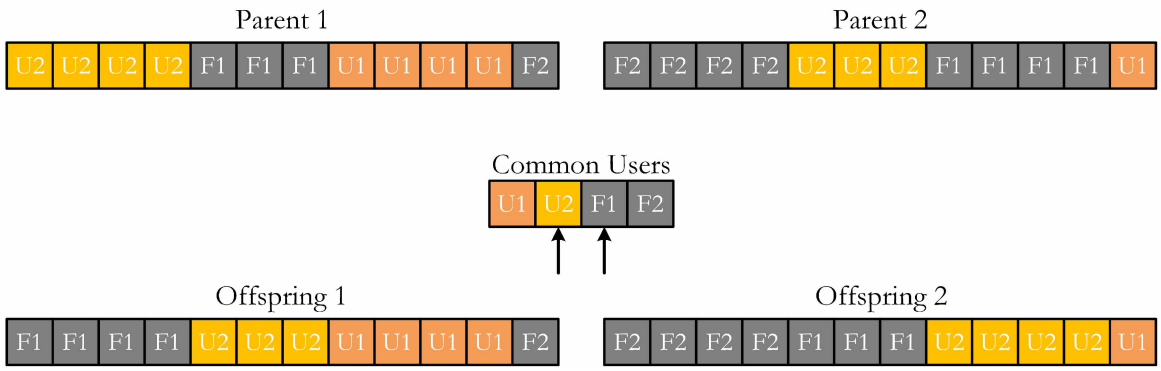


Figure 5.8: Crossover operation considering discontinuities and fake users.

Mutation

The mutation operation is another part of the recombination phase. This operation performs small changes in the chromosome, diversifies the population and prevents the solutions of being trapped in a local optimum. As mentioned for crossover operation, we developed a mutation strategy that respects the contiguity constraint. Figure 5.9 shows an example of the mutation operation.

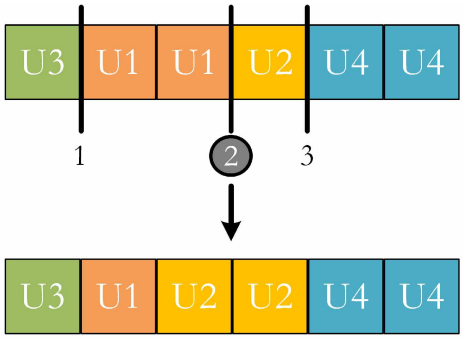


Figure 5.9: Mutation operation.

Consider the chromosome illustrated in Figure 5.9. The first step is the identification of the boundaries between the RBs granted to different users. These boundaries are stored in a vector. In the example, we can identify three boundaries. In the next step, we randomly choose one of these boundaries. In the example, boundary number 2 was chosen. In the third step, we "flip a coin" to select one of the users that form boundary number 2. The winner takes the RB of the other user. In Figure 5.9, user U2 won the challenge and took the RB that was formerly granted to user U1.

For the case when we have the presence of discontinuities in the chromosome, the algorithm must identify if a particular boundary is also a discontinuity. Only boundaries that are not discontinuities are considered in the mutation operation. Otherwise, we would break the contiguity in the RB assignment. In the example of Figure 5.10, only

boundary number 1 is suitable for the mutation operation.

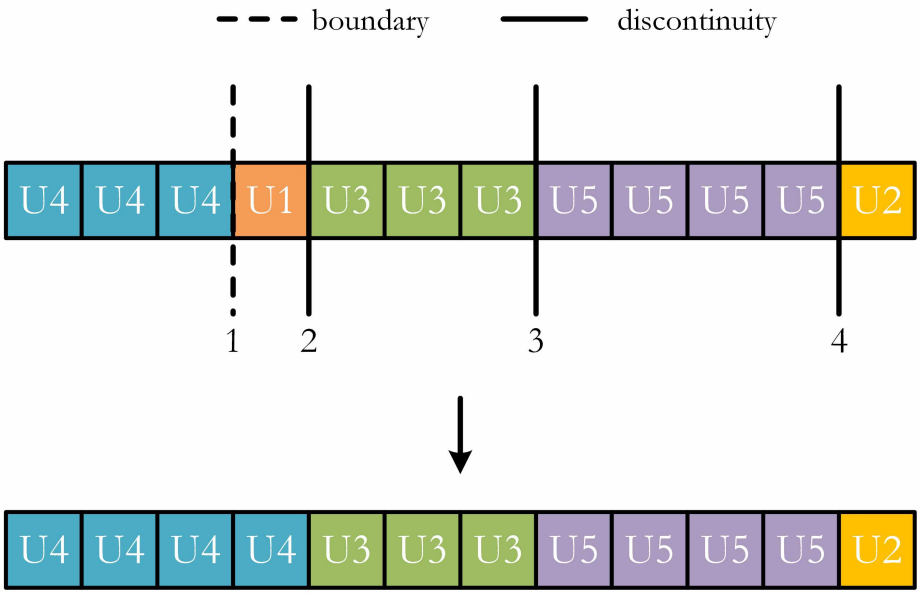


Figure 5.10: Mutation operation considering discontinuities.

Termination

As shown in Figure 5.2, we have to set a termination criterion to end the algorithm. A common end criterion is to define a number of generations (iterations) to terminate the algorithm.

5.3 Summary

In this chapter, we proposed a new Three-Step GA-based scheduling algorithm for the LTE uplink. In this sense, while designing our scheduler, we introduced new methods to perform the genetic operations. These methods always considered the contiguity constraint for the RB assignment and the existence of discontinuities in the resource grid caused by HARQ retransmissions as well.

It is worth saying that of all operations presented in this chapter, only the *initialization* can not take advantage of the improvements introduced by Release 10 concerning the two non contiguous groups of RBs. The other operations remains the same and can work in both scenarios.

In the next chapter, we will evaluate the proposed algorithm and compare its performance with the performance of the algorithms presented in the Chapter 4 to verify

the relevance of our proposal.

"Do you know what we call opinion in the absence of evidence? We call it prejudice."

- Michael Crichton, *State of Fear*

6

Performance Evaluation

PERFORMANCE EVALUATION is a fundamental step to verify how good a model or an algorithm is. In this sense, this chapter presents a performance evaluation of the algorithm proposed in Chapter 5. We start by presenting the simulation environment where the evaluation was performed. Then, the parameters of the simulations are stated. Finally, simulation results are presented.

6.1 Simulation Environment

A performance evaluation can be conducted using several techniques and approaches, depending on the subject of study. One of the main approaches is the use of a real system where an algorithm can be evaluated. Generally, this approach is very expensive and complex to manage. This is the case for the LTE system. As presented in Chapter 2, the LTE network is a huge and complex system composed of several network elements, interfaces and protocols. The evaluation of a scheduling algorithm in a real system would require the use of several equipments that are expensive even for network operators. Besides, this scenario would also require a special environment to control the interference of electromagnetic waves emitted by the several services using the shared spectrum.

In this sense, to avoid some of these problems, performance evaluation can also be conducted by means of *models*. System modeling refers to an act of representing an actual system in a simple way [37]. This is an important approach that reduces the costs of the evaluation, while offering results close to the ones obtained with an evaluation in a real scenario. The closeness of the results will depend on the model used. Robust models offer better results than simplistic models, in exchange of a more complex implementation. Thus, it is necessary to adjust the model assumptions according to the required accuracy.

Traditionally, there are two modeling approaches:

- **Analytical approach:** makes use of mathematical tools, such as queuing and probability theories, to describe the system behavior. This approach is preferable when the system is simple and small, since, in this case, the model tends to be mathematically tractable.
- **Simulation approach:** the simulation approach is more indicated to complex systems. This approach can support a robust model with many details of the real system. However, the greater the number of details, the greater the computational effort to run the simulation.

Since the LTE network is a complex system, the simulation approach is more appropriated to conduct a performance evaluation. Once the evaluation approach was chosen, it is necessary to define the software, techniques and mathematical tools required to run the simulations.

LTE has been a field of intense research. However, so far, 3GPP has not launched an official simulation environment for the LTE network, but only some reference system deployments reports to use for different system evaluations [38]. This fact promoted the rising of a great number of simulation tools for the LTE network, distributed into a set of categories: private, commercial, open source, etc.

Among these several solutions, we searched for a simulation tool with the following features:

- **Open source:** most of open source software are free of costs. Commercial network simulators usually present a high cost per license. Despite of the slower development speed, open source software generally allow customization of the code for particular purposes.
- **Reliable models:** the simulation tool must present reliable models to assure the validity of the results.

- **Efficient implementation:** to simulate a wide range of scenarios, the simulation tool should present a modern implementation to use the hardware resources in an effective manner.
- **Good documentation:** a well documented software is of fundamental importance to facilitate the use of the tool and to promote contributions by the users.

Considering the features mentioned above, the Network Simulator 3 (ns-3) was chosen as the simulation environment for the performance evaluation.

ns-3 is a discrete-event network simulator, targeted primarily for research and educational use. It is a free software, licensed under the GNU GPLv2 license, and is publicly available for research, development, and use [39]. This software is a general purpose network simulator, i.e., it supports simulations for several types of networks. This is possible because the development of ns-3 is based on modules as shown in Figure 6.1. The core and modules of ns-3 are implemented in C++. From Figure 6.1, one can see that the core of the simulator is responsible to provide functionalities that are common to several types of networks, such as events, time arithmetic and random variables. The network module is responsible for implementing the packet structure of the network. The helper module facilitates the process of building a simulation script. These scripts are also written in C++ and specify the parameters of the simulation (topology, air interface, simulation time, etc). It is worth saying that ns-3 also supports scripts wrote in Python. Similarly, Figure 6.1 illustrates other modules responsible for specific functionalities.

In what concerns the LTE network, ns-3 presents a module called LTE-EPC Network simulAtor (LENA) [41]. This module has been developed by the researchers of the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), aiming to facilitate the design and performance evaluation of downlink and uplink schedulers, RRM algorithms, among other functionalities. LENA module offers the basics features of the LTE Release 8 and it presents a good documentation, which allows the extension of the module for specific purposes.

In this context, we implemented and integrated the Three-Step GA-based algorithm into the ns-3 structure. We also implemented the RPQoS and RME algorithms, described in Chapter 4. The RR algorithm is already part of the ns-3 official release. In the next section, the simulation setup is detailed.

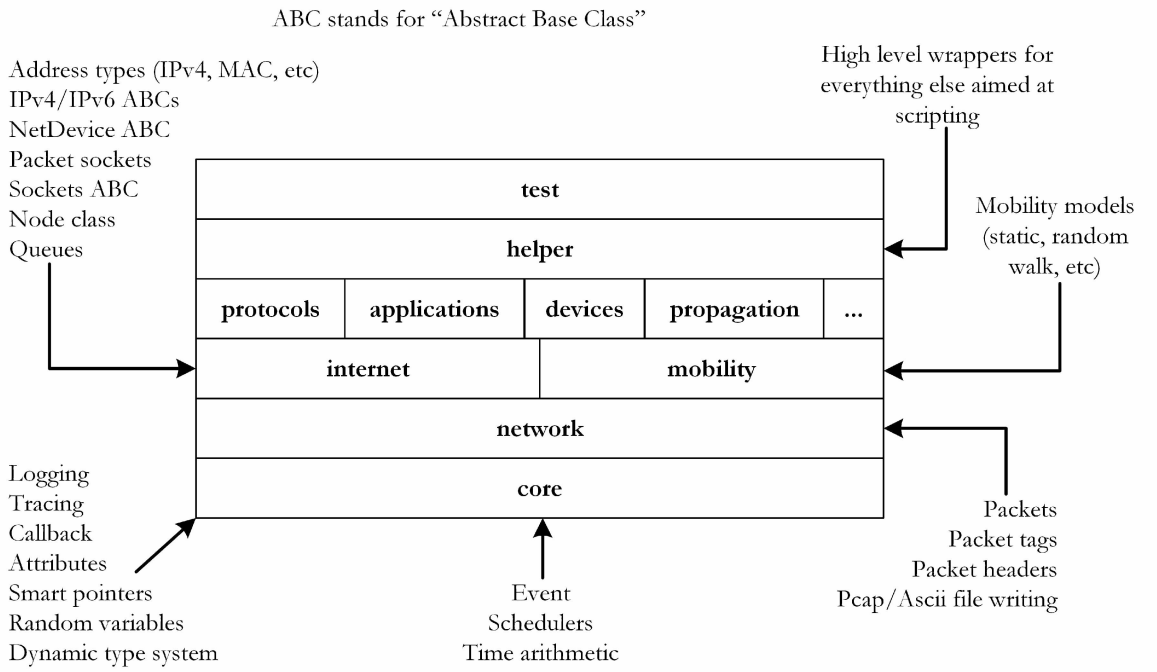


Figure 6.1: ns-3 software organization (Adapted from [40]).

6.2 Simulation Setup

In this section the simulation setup used in the evaluation is presented. We used a typical outdoor scenario to evaluate the algorithm. Figure 6.2 shows the topology used in the evaluation.

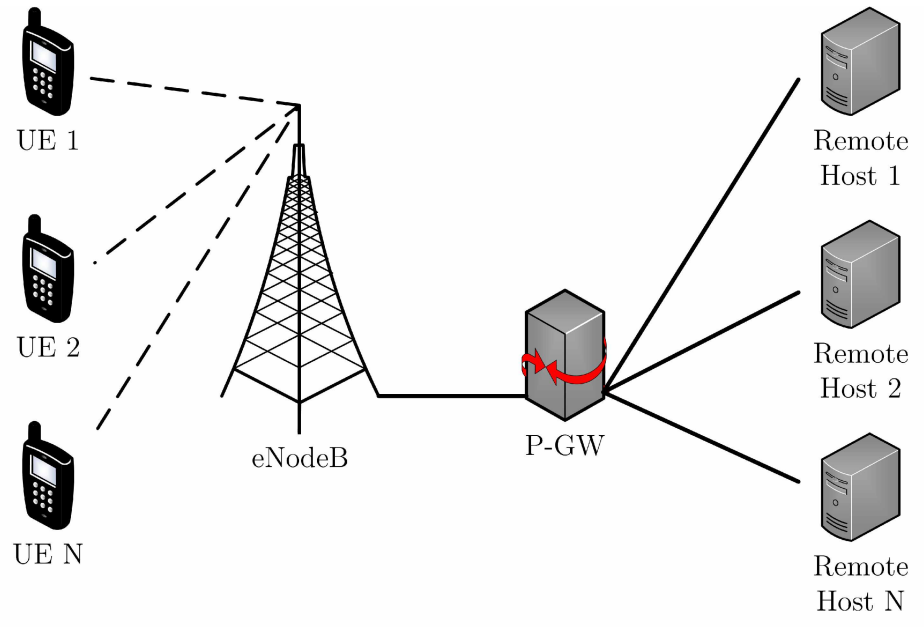


Figure 6.2: Simulation topology.

The topology of Figure 6.2 is based onto a single macro cell scenario. This means that the simulation scenario considered only one eNB with an isotropic antenna, and therefore, only one sector. In this scenario, inter-cell interference was not considered in the simulations. However, none of the scheduling algorithms evaluated in this research makes use of the inter-cell interference values to allocate the resources of the base station. In this sense, the inter-cell interference becomes a constant in the comparison.

The use of a single macro-cell simplifies the simulation scenario, since we don't have to answer questions that are not relevant for this type of comparison such as: How many cells should be used? What type of antenna should be employed? Will there be users in all cells or only in the center one? How to deploy the users among the sectors of the cell? Will the handover be allowed? etc. It is worth saying that the evaluation considered the interference caused by the transmission of one UE into the transmissions of the other UEs inside the cell.

The macro cell was populated with a set of UEs attached to the eNB. The deployment of the UEs inside the cell was based on the air interface and mobility models offered by ns-3. In the simulations, the UEs could move according to the Steady State Random Waypoint mobility model [42] [43]. In this model, the UE starts at a pause mode. The interval that they remain static is controlled by a random variable with parameters of minimum and maximum pause interval defined in the simulation script. After this pause interval, the model chooses a new destination for the mobile. This new destination is a position inside a square area, whose size was previously defined in the simulation script. Then, the UE starts to move at a speed controlled by a random variable with parameters of minimum and maximum speed also defined in simulation script. Once it has arrived at the destination, it enters in pause mode and the cycle is restarted.

As said before, the UEs are attached to the eNB through the air interface. The air interface is characterized by a set of phenomena, such as path loss, fading, interference, etc. ns-3 provides a set of models to mimic these phenomena, according to the scenario of evaluation. For mobile networks outdoor scenarios, it is common to use the COST 231 model to calculate the path losses [44]. In this sense, we used an empirical approach and performed some simulations to understand how severe the loss calculated by this model was, considering the presence of the fading phenomena, and then, define how far the UEs could be deployed and still be attached to the eNB. From the SINR measured in the uplink channel, the square area was set to 424x424 meters. In this area, the longest distance between a UE and the base station would be 300 meters, which correspond to a MCS equals to 2, as depicted in Figure 6.3. Typically, the UEs experienced MCS values from 2 to 28. Figure 6.4 shows the average Modulation and Coding Scheme (MCS)

levels and the Cumulative Distribution Function (CDF) values experienced by the users in the simulations. This figure indicates that all scheduling algorithms were evaluated with very similar channel conditions, allowing a fair comparison.

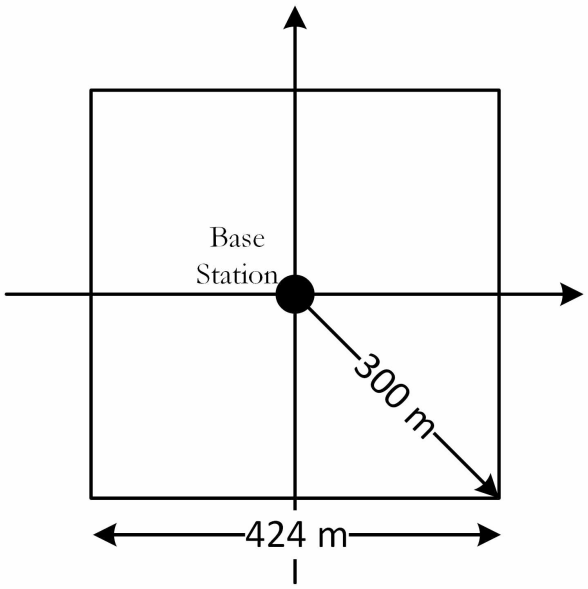
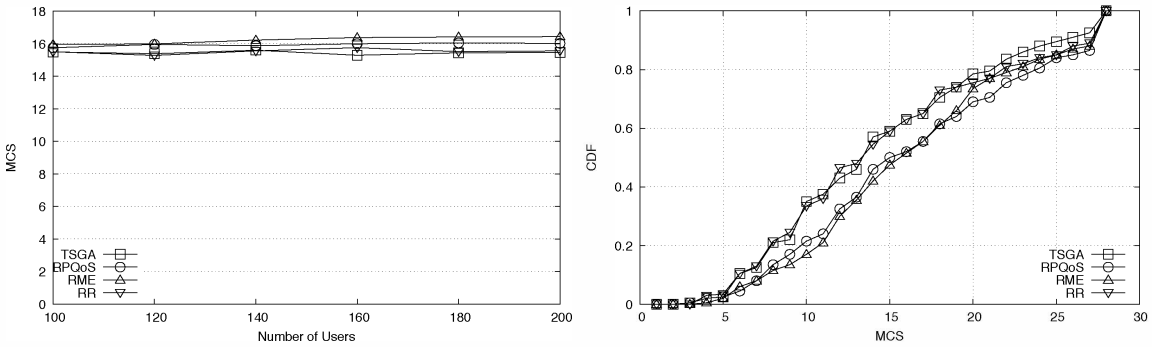


Figure 6.3: Deployment area.



(a) Average MCS. (b) CDF of MCS (200 users in the cell).

Figure 6.4: MCS: average and CDF values.

ns-3 also provides a model for the fading phenomenon. The fading model is based on trace files with previously calculated values. This approach was chosen to reduce the complexity of the model in run time. Then, ns-3 offers a group of trace files with values of fading for specific scenarios, such as Extended Pedestrian A (EPA) and Extended Vehicular A (EVA) models. We performed preliminary tests to verify the differences of using these two models. Results showed that they performed similarly. This occurred because the scheduling algorithms allocate the resources according to the users's channel quality. However, they perform this allocation considering the status of all users. In this sense, when we observe the performance indicators for the cell as a whole, there is

no significant difference between EPA and EVA scenarios. In this sense, we chose the fading traces for a scenario with a pedestrian walking at 3 km/h (EPA), since it is a more realistic scenario than the one with users moving at 60 km/h in such small area defined for the evaluation. From this information, it is also possible to define the speed values of the mobility model.

From Figure 6.2 one can see that the eNB was connected via the P-GW to remote hosts. For each UE, a remote peer was connected to the P-GW via a separated point to point link with over provisioned bandwidth. Then, applications were installed in the peers according to a client-server architecture, i.e., the remote host requests packets from the UE, which in turn sends them to the remote host.

The scenario of evaluation considers a mixed traffic environment. In this sense, three applications were deployed in the UEs: Video, VoIP and FTP. As said in section 5.2.3, in this thesis, there is a particular interest in Video Chat application because of the challenges that this kind of application brings to the uplink channel. VoIP and FTP applications were included in the evaluation to verify how the schedulers behave when different types of applications require resources from the network. The FTP traffic represents all types of applications that belong to Non-GBR classes, while the VoIP application represents the GBR applications. VoIP and FTP applications were modeled by a traffic source that sends packets of constant size at regular intervals, characterizing a particular bit rate. Table 6.1 describes the details of the traffic sources used in the evaluation.

Table 6.1: Traffic Sources.

Application	**Bit Rate (kbps)**	**QCI**	**Packet Delay Budget (ms)**
VoIP	8	1	100
Video	128	2	150
FTP	512	9	300

The video application used is based on EvalVid [45]. EvalVid is a framework and tool-set for evaluation of the quality of video transmitted over a real or simulated communication network. It is targeted for researchers who want to evaluate their network designs or setups in terms of user perceived video quality. From the original framework, several modules were created to use EvalVid in different simulators. Officially, ns-3 does not offer support for EvalVid. However, there is an unofficial module developed by the community of users that is not part of the official release yet. The ns-3 EvalVid module can be found in [46].

Then, EvalVid was setup to model a Video Chat transmission. The first step is

the choice of the video sequence. For the current evaluation, the well-known video sequence *akiyo* was chosen. This video sequence presents a news reporter talking and it was chosen since it has similar characteristics to Video Chat, i.e., a low motion video sequence. Hence, this video sequence was encoded using parameters values close to the ones used in real Video Chat transmission, as depicted in Table 6.2.

Table 6.2: EvalVid Simulation Parameters.

Video sequence	akiyo (300 frames)
Video resolution	QCIF (176 x 144)
Video info	Bit rate: 128 kbps Frame rate: 25 fps Group of Pictures: 30 MTU: 1460 Bytes Encoder: ffmpeg

In the beginning of the simulation, all UEs get connected to the eNB. Then, in the interval between one and two seconds, each user randomly begins to require resources from the network. The akiyo video sequence presents duration of 12 seconds. Therefore, the VoIP and FTP applications were adjusted to require resources from the network in this interval. The total simulation time was set to 30 seconds. This was necessary to assure that all packets had enough time to reach their destination. Each scenario of evaluation was repeated 50 times and the results present the confidence interval of 95%.

Table 6.3 summarizes the parameters used in the simulations and their values. In what concerns the GA parameters, the values were chosen in an empirical manner, considering the tradeoff between performance and complexity. For tournament and elitism size, common values used in the literature were chosen. In these preliminary tests, population size was the main factor to delay the execution time. Then, a small value for this parameter was selected. This was possible since Step Two allows only five users to participate in Step Three of the proposed algorithm. The preliminary tests also showed that after about five generations, the diversity of the population was decreased and there was no reason to continue the genetic operations. Since population size and generations were set to small values, crossover and mutation rates should be adjusted to high values, allowing the algorithm to evaluate a wider range of possible solutions. The elitism strategy avoided the best solution to be destroyed by this high rates of the recombination phase. Further details about the simulation parameters can be found in the ns-3 documentation [39].

Table 6.3: Simulation Parameters.

eNB antenna model	Isotropic antenna model
eNB TX Power	46 dBm
UE TX Power	23 dBm
Bandwidth	50 RBs (10MHz)
AMC scheme	PiroEW2010
RLC mode	UM (buffer size: 10 MB)
Pathloss model	COST 231
Fading loss model	Pedestrian EPA 3 km/h
User mobility model	Steady state random waypoint Min speed: 0.8 m/s / Max speed: 0.83 m/s Min pause: 0 s / Max pause: 0.1 s Rectangle: 424m x 424m
Simulation time	30 seconds
Simulation runs	50
Confidence interval	95%
GA population size	10
GA tournament size	2
GA elitism size	1
GA crossover rate	0.95
GA mutation rate	0.95
GA generations	5
Traffic Distribution	[FTP:VoIP:Video] [0.5:0.25:0.25]

6.3 Simulation Results

From the topology and parameters values detailed in the previous section, scenarios of evaluation were built with distinct number of UEs in the macro-cell and the performance of the scheduling algorithms under evaluation was analyzed, considering some network performance indicators.

6.3.1 Throughput

Figure 6.5 shows the aggregated cell throughput found in the simulation. As described in Table 6.3, the scenario of evaluation considered a traffic distribution where 50% of users were running the FTP application, 25% Video Chat and the last 25% VoIP.

The RR algorithm allocates the resources by giving the same opportunities to all users. Since there are more users running the FTP application, which presents the highest bit rate, the RR algorithm presented the highest aggregated cell throughput. On the other hand, TSGA has a compromise with GBR applications. TSGA tries to assure the QoS requirements of these applications and this may mean to sacrifice the Non-GBR applications. As a consequence, TSGA presented a lower aggregated cell throughput when compared with QoS-unaware schedulers. However, when compared with RPQoS, which is a QoS-aware scheduler, TSGA presented a superior performance.

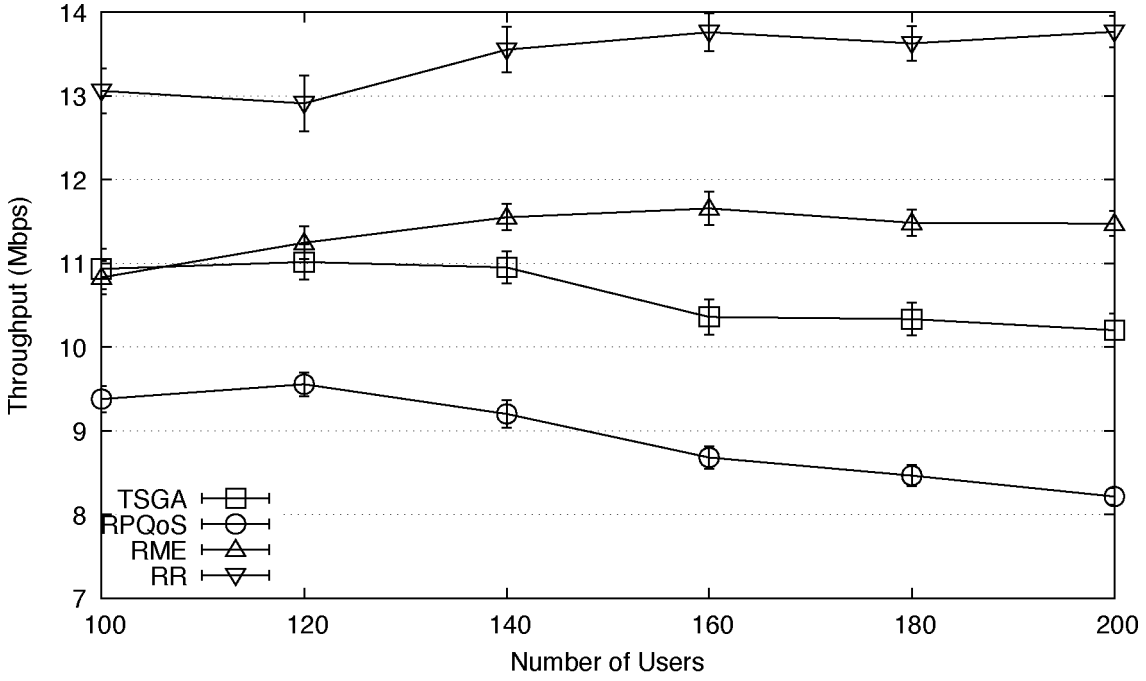


Figure 6.5: Aggregated cell throughput.

Figure 6.6 shows the aggregated cell throughput considering only the users running the Video application. It is worth saying that, despite the figure shows only the performance for video application users, the scenario is the same of Figure 6.5, i.e., users of FTP and VoIP application were also running in the cell. Figure 6.6 illustrates the superior performance of QoS-aware schedulers. This figure also shows that TSGA presented a vigorous performance when compared with RPQoS, mainly when the cell is crowded and the resources are more scarce. Considering 200 users in the cell (50 of them video users), RPQoS is in saturation zone, while TSGA still can keep the aggregated cell throughput growth. This fact indicates that GA was able to offer a better mapping than the RP strategy for the RBs and users in the evaluation. This more efficient mapping improved the spectral efficiency of the cell.

In Figures 6.5 and 6.6, it is also possible to see that RR performed better than RME. Despite the fact that RME is based on a PF metric, which is superior than

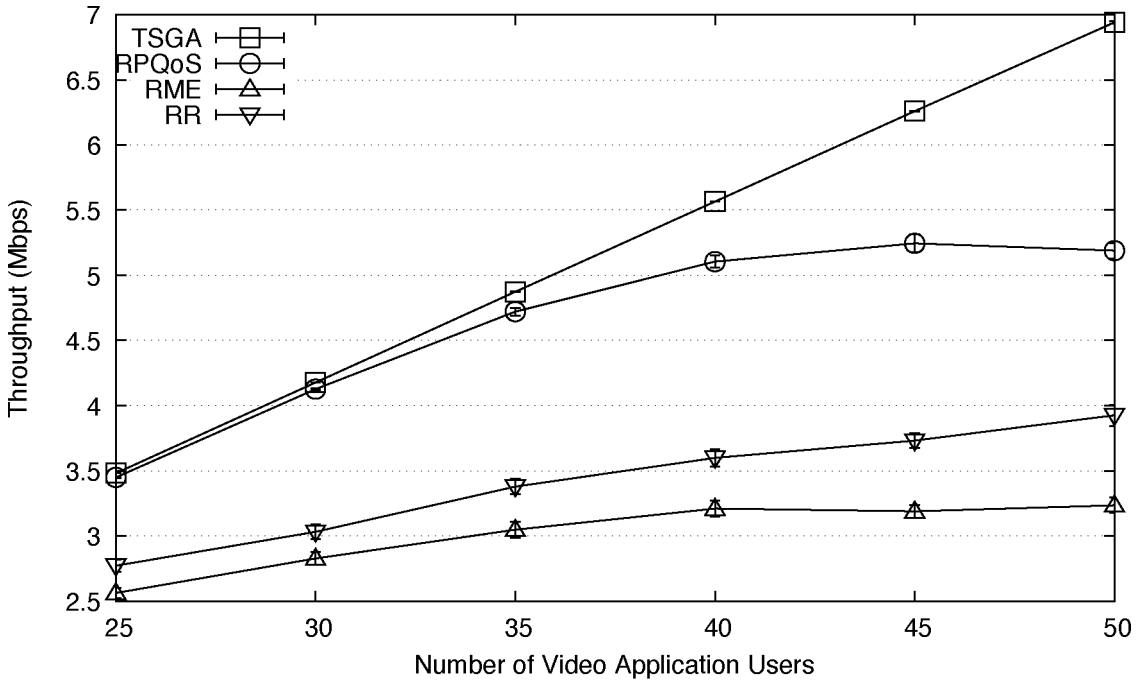


Figure 6.6: Aggregated cell throughput for users running the Video application.

the RR metric for throughput, the RME allocation strategy does not perform well in mixed traffic environments. We believe that RME was not able to efficiently adjust the allocation process when there were users requiring resources at different bit rates. This fact was also concluded by the authors of [34].

Continuing the analysis of the network when it is crowded, and therefore, more challenging, Figures 6.7 and 6.8 show the CDF of the throughput for FTP and Video applications, respectively, considering 200 users in the cell. In Figure 6.7, one can see that RR and RME presented the highest throughput, in consonance with Figure 6.5. It is also possible to see that the QoS-aware schedulers presented lower throughput, but without starvation for FTP users.

In Figure 6.8, the Video traffic belongs to a GBR class, and therefore, it presents a minimum bit rate to work properly. As stated in Table 6.1, the Video Chat application used in the simulation is based on a traffic with bit rate of 128 kbps. Considering the overhead, each user of this application requires a bit rate of about 140 kbps from the network. In this sense, Figure 6.8 indicates that only TSGA was able to meet this requirement, when the cell has 200 users.

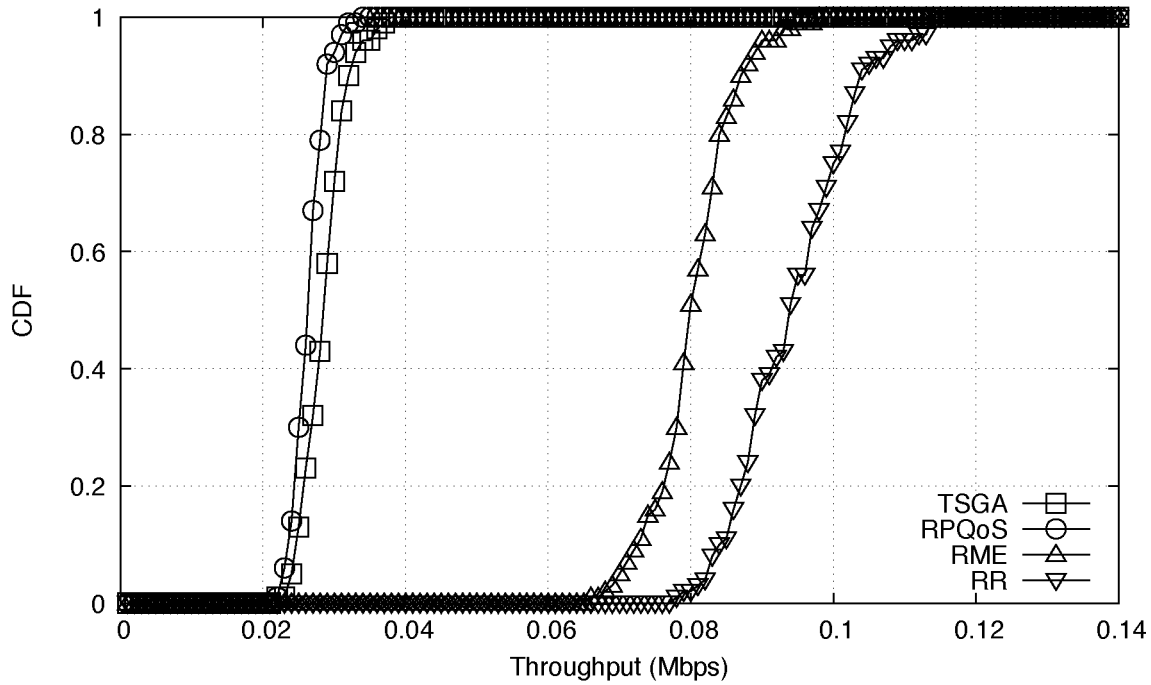


Figure 6.7: CDF of throughput for users running the FTP application, considering 200 users in the cell.

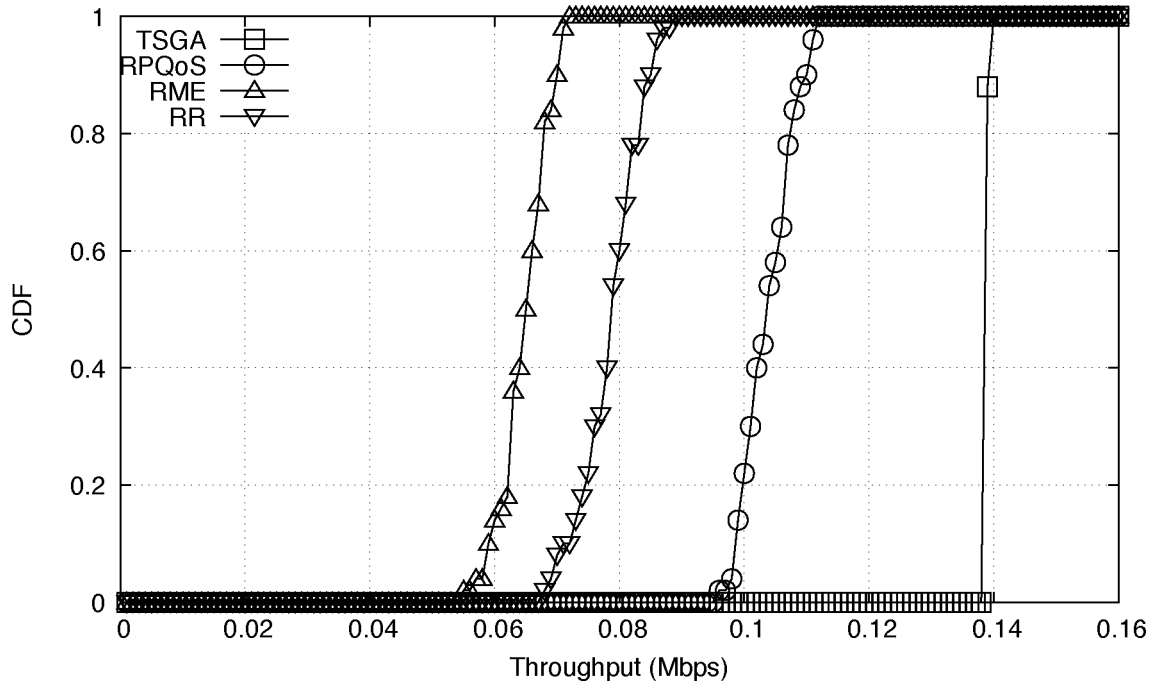


Figure 6.8: CDF of throughput for users running the Video application, considering 200 users in the cell.

6.3.2 Packet Delay

Table 6.1 indicates the packet delay budget defined for the applications used in the simulation. From these values, we have implemented a routine in ns-3 to discard packets in the UE's PDCP buffer that reach the packet delay budget. In this sense, Figures 6.9, 6.10 and 6.11 show the average delay for FTP, Video and VoIP traffics, respectively.

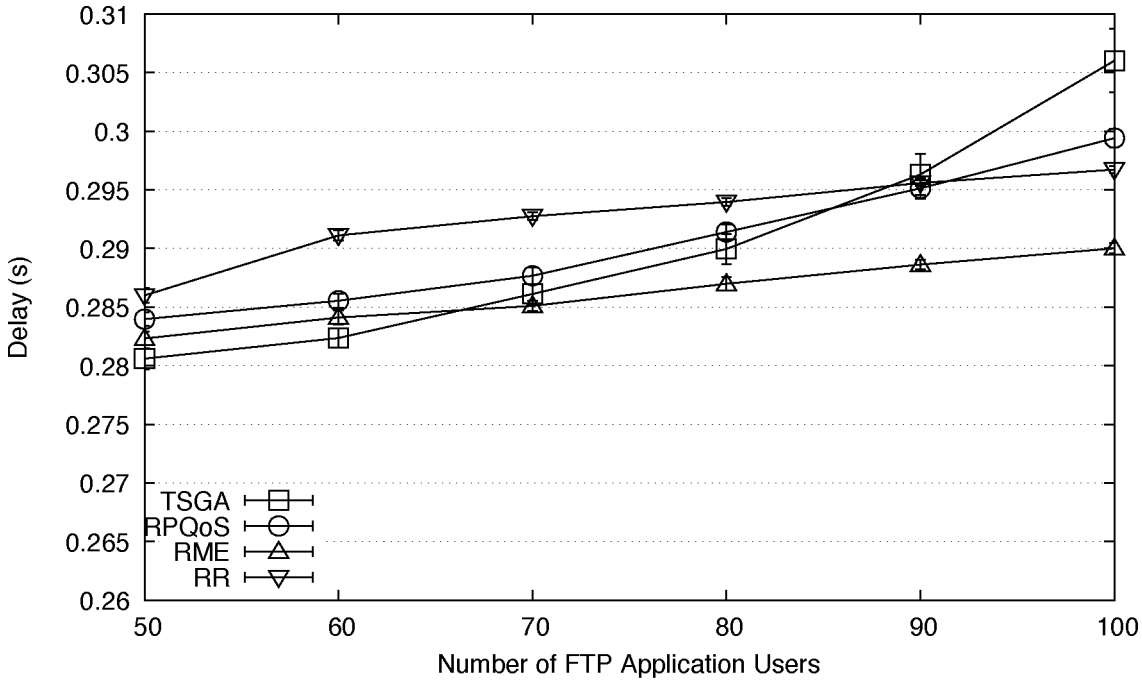


Figure 6.9: Average delay for users running the FTP application.

From the figures, it is possible to note that until 160 users in the cell, all schedulers met the delay requirement defined for the applications. However, when 200 users were placed in the cell, some schedulers did not meet this criterion. This happened because we applied a simplistic approach for the discard packet routine. Hence, the routine only discards full packets in the buffer, i.e., if a packet is under transmission and reaches the maximum delay, it will not be discarded by our routine. This approach was chosen since we used the sequence number field of PDCP header to track de delay of the packets. When the Protocol Data Unit (PDU) is not fully stored in the buffer, the PDCP header is not available and it will be need another strategy to correctly track these bytes through all layers of the protocol stack. As a result, a packet under transmission that is not discarded will be received with a packet delay greater than the packet delay budget, as depicted in Figures 6.9 and 6.10.

It is worth saying that the average delay statistics should not be individually analyzed. It is necessary to consider the impact of the discarding process in the Packet Loss Ratio (PLR).

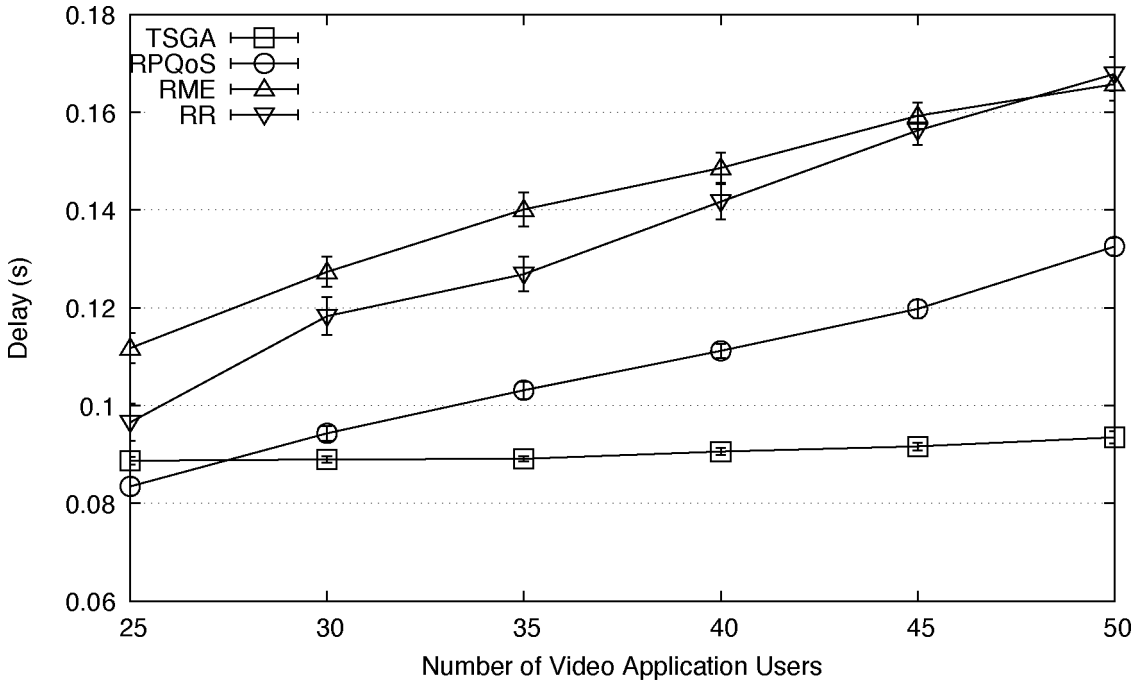


Figure 6.10: Average delay for users running the Video application.

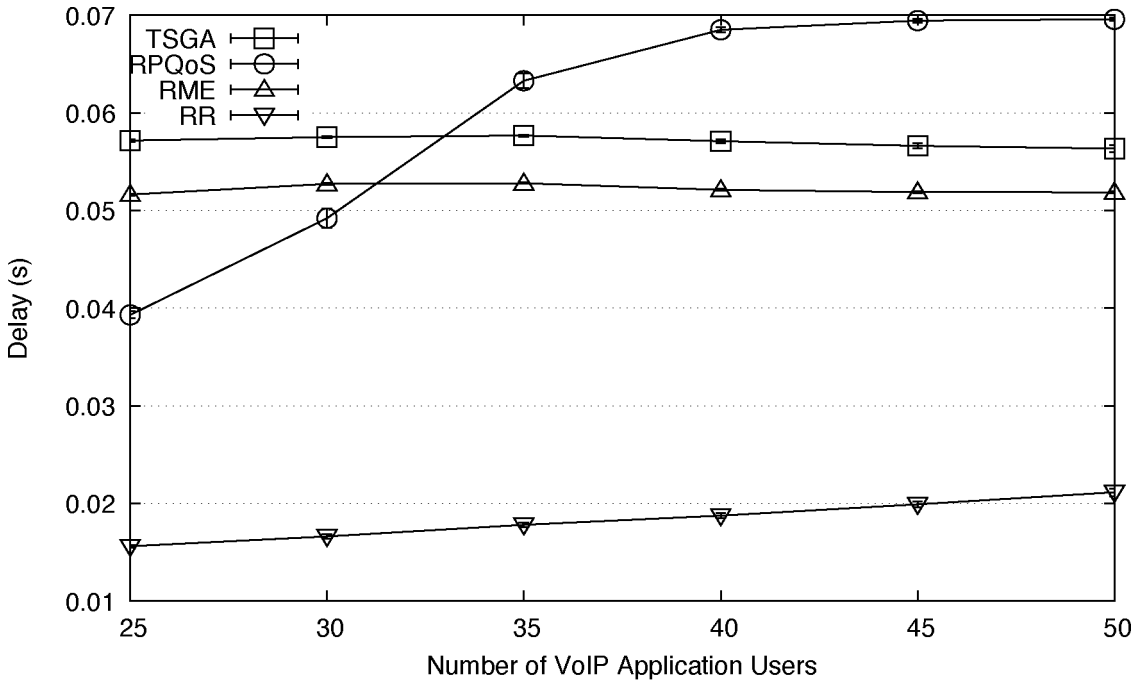


Figure 6.11: Average delay for users running the VoIP application.

6.3.3 Packet Loss Ratio

The PLR was calculated from the number of packets lost due to reaching the packet delay budget. The RLC buffers in the UEs were set to 10 MB to assure that the losses

were caused by the discarding process and not affected by buffer overflows.

Figure 6.12 shows the average PLR for users running the FTP application. As expected, the QoS-aware schedulers sacrifice this traffic to meet the QoS requirements of GBR traffics, and therefore, they presented the highest PLRs. It is worth saying that TSGA presented a lower PLR when compared with RPQoS.

Figures 6.13 and 6.14 show the average PLR for users running Video and VoIP applications, respectively. In both figures, TSGA presented very low PLR levels. These figures also explicit the poor performance of RME scheduler. The performance of this algorithm degrades significantly in the presence of mixed traffic environments, as said before.

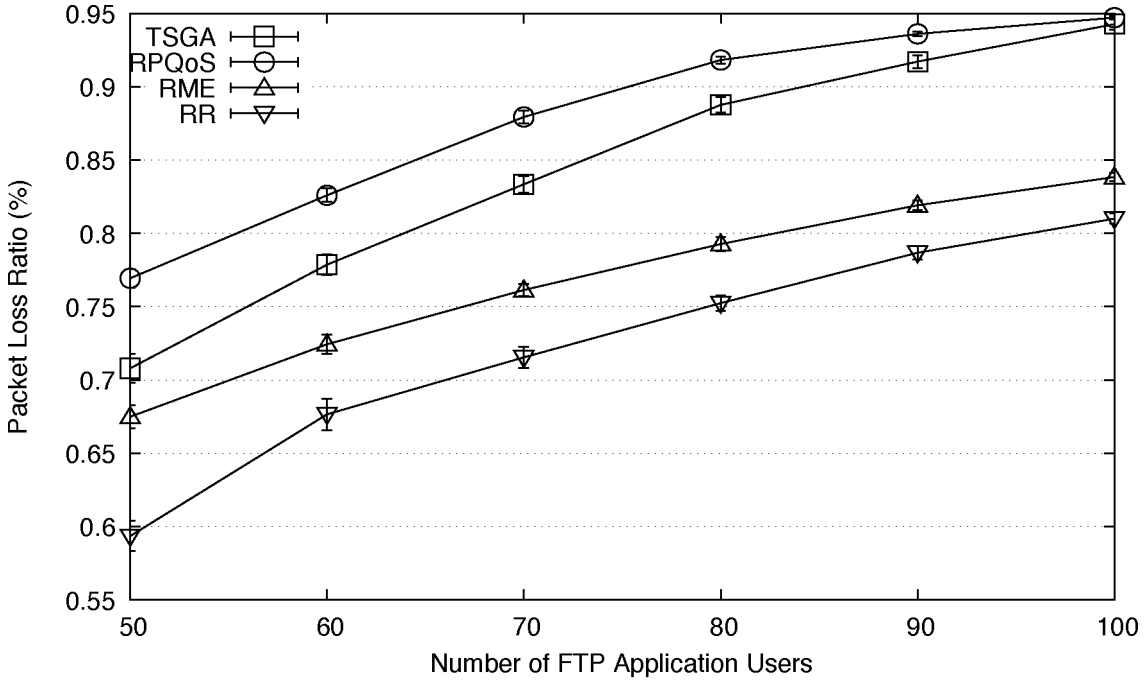


Figure 6.12: Average PLR for users running the FTP application.

In this sense, TSGA presented the best performance, meeting the packet delay budget, while keeping the PLR at satisfactory levels. We believe that the cause of this performance is the three step approach: Step One filters the most urgent users with fair weights between GBR and Non-GBR users; Step Two allocates the users with packets very close to the packet delay budget, reducing the number of packets considered lost; finally, Step Three efficiently allocates the resources, reducing the probability of a packet to be lost.

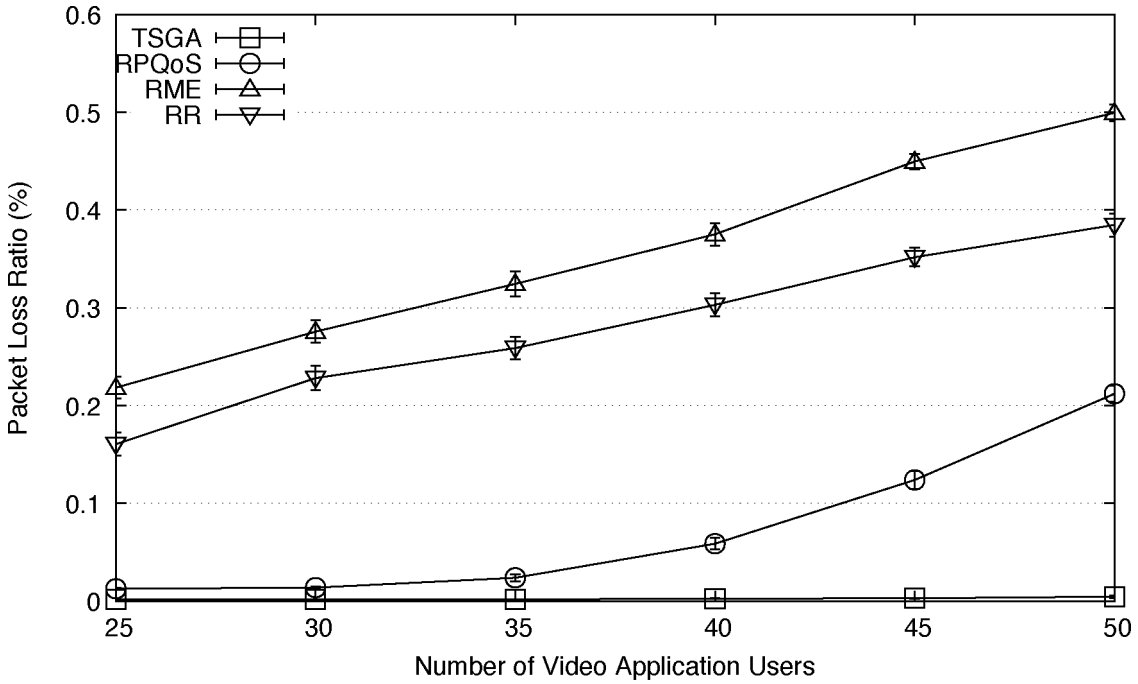


Figure 6.13: Average PLR for users running the Video application.

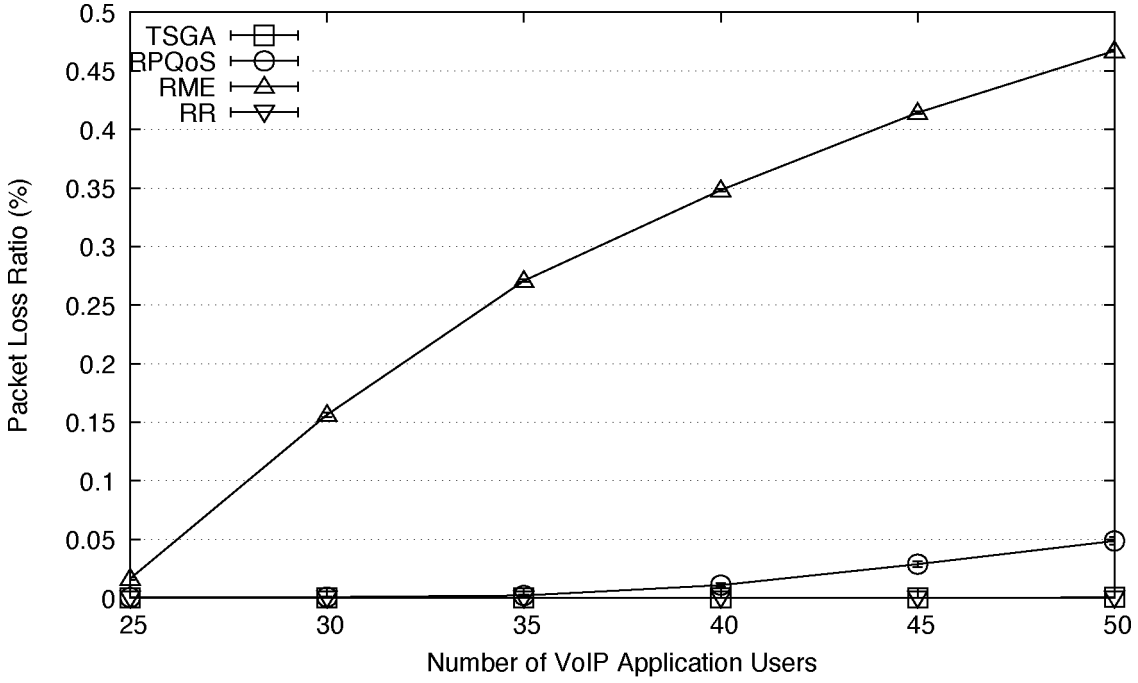


Figure 6.14: Average PLR for users running the VoIP application.

6.3.4 Throughput Fairness Index

In what concerns throughput fairness, we applied the well-known Jain Fairness Index [47]. We considered two types of fairness: inter-class and intra-class fairness. In the

inter-class fairness the goal is to measure for how many users the network could assure the bit rate they required. Therefore, to achieve the inter-class fairness between different QoS classes, the average throughput of the UEs was normalized with respect to the bit rate defined in Table 6.1. Figure 6.15 shows the results. As expected, RR and RME presented better inter-class fairness levels since they give more attention to FTP users when compared with TSGA and RPQoS. The FTP users are the majority in the cell and when the QoS-aware schedulers sacrifice their throughput to meet the requirements of the GBR applications, the inter-class fairness index of these schedulers is compromised. Figure 6.15 shows once more that TSGA meets the QoS requirements of GBR applications with less impact in the Non-GBR applications, when compared with RPQoS.

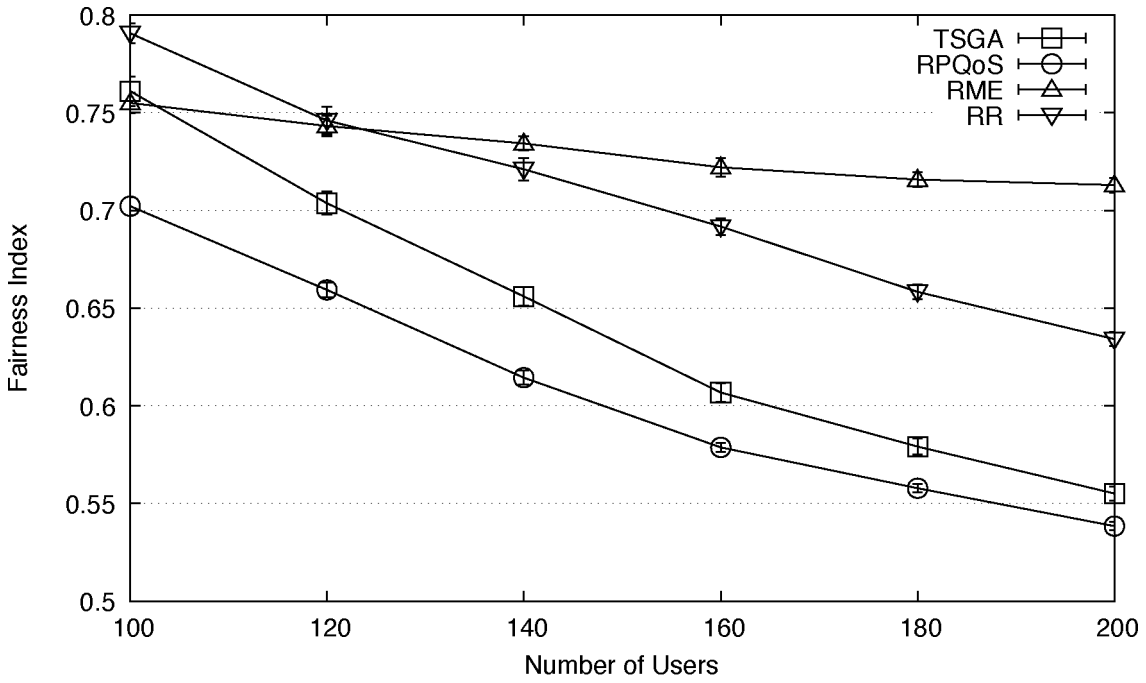


Figure 6.15: Inter-class fairness index.

The intra-class fairness index tries to indicate how fair the scheduler is while allocating resources among users of the same QoS class. Figure 6.16 shows the intra-class fairness index for users running the FTP application. One can see that TSGA presented the best intra-class fairness index levels among the schedulers evaluated, until 80 FTP users in the cell. Then, in order to attend the GBR applications, TSGA starts to present difficulties to keep a high fairness index for Non-GBR users. On the other hand, the good performance for fewer users was possible due to our fitness function described earlier in Equation 5.4. This fitness function tries to minimize the error committed by the scheduler while allocating the resources. In this sense, as the FTP users demand the same resources from the network, this equation tries to grant the same amount of

resources for each one of them.

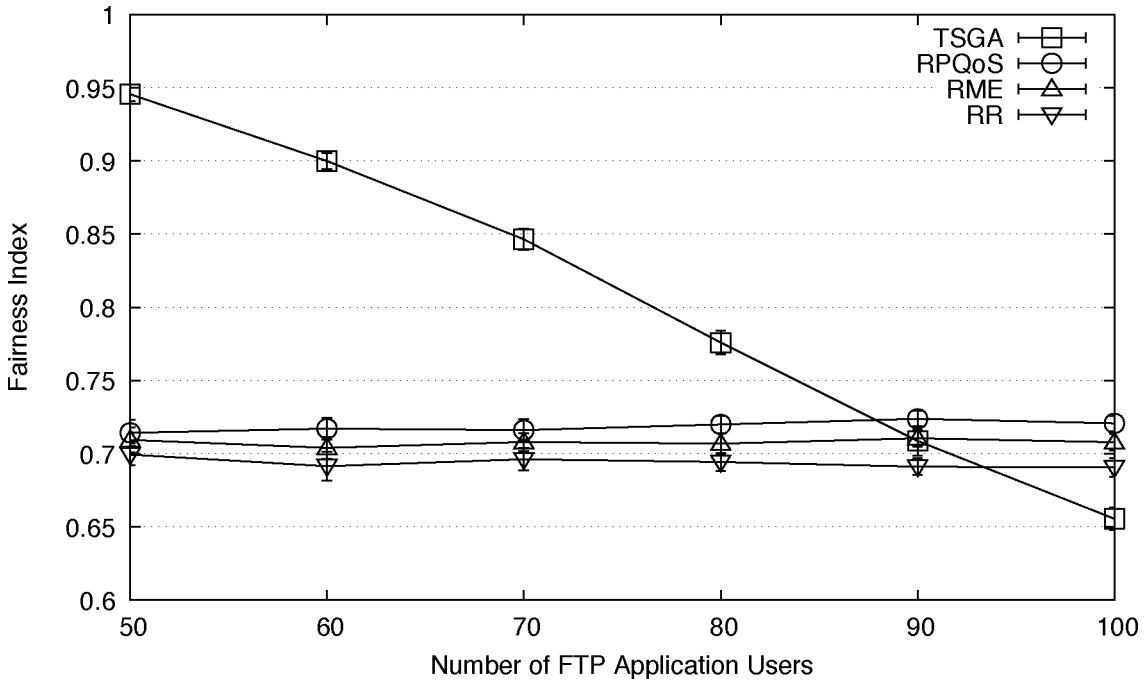


Figure 6.16: Intra-class fairness index for users running the FTP application.

Figure 6.17 indicates the intra-class fairness index for users running the Video application. It is possible to note that TSGA, once more, presented the best performance. Despite of the cell load and user's channel quality, TSGA tries to equally serve all users. RPQoS presented similar performance with few users in the cell, but it was not able to keep the intra-class fairness index. RR and RME were not able to follow the performance of the QoS-aware schedulers.

Finally, Figure 6.18 shows the intra-class fairness index for users running the VoIP application. Since the VoIP traffic presents a low bit rate, all schedulers performed similarly.

6.3.5 PSNR

The Peak Signal-to-Noise Ratio (PSNR) is one of the most employed objective metrics in evaluation of video transmission quality. Hence, Figure 6.19 indicates the average PSNR found in the evaluation. It must be noted, that PSNR cannot be calculated if two images are binary equivalent. This is due the fact that the mean square error would be zero and thus, the PSNR could not be calculated. Usually this is solved by calculating the PSNR between the original raw video file before the encoding process

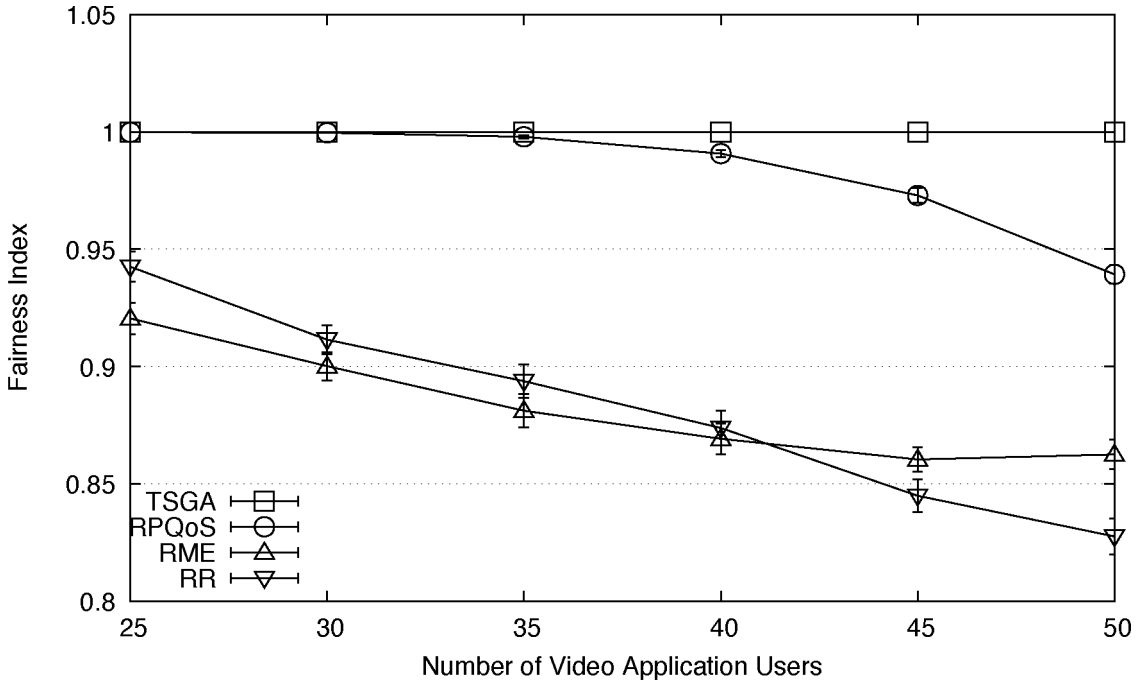


Figure 6.17: Intra-class fairness index for users running the Video application.

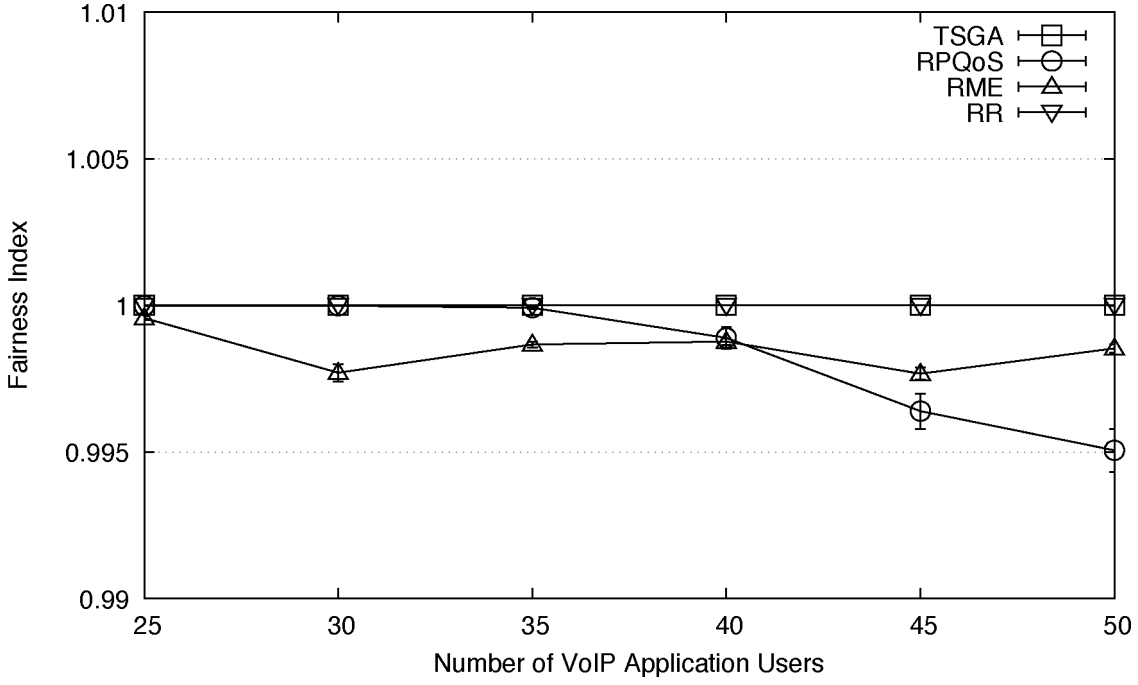


Figure 6.18: Intra-class fairness index for users running the VoIP application.

and the received video. This assures that there will be always a difference between the two raw images, since all modern video codecs are lossy.

As expected, the PSNR followed the tendency of the previous performance indicators and confirmed the superiority of TSGA over its competitors. TSGA was able to keep

the PSNR almost constant as the number of users in the cell was increased.

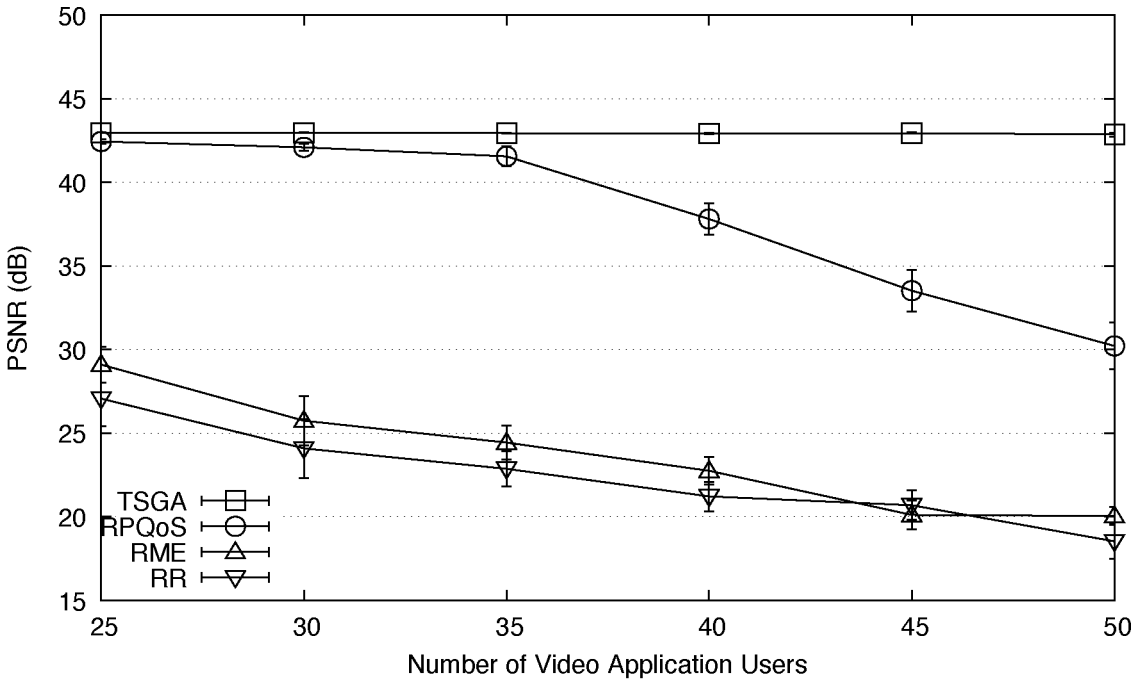


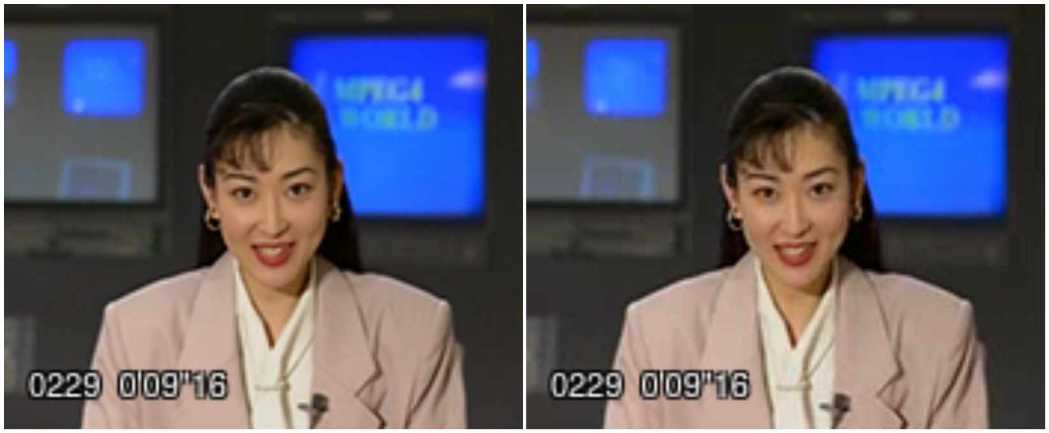
Figure 6.19: PSNR for users running the Video application.

To facilitate the interpretation of Figure 6.19, one should consider the frames depicted in Figure 6.20. In this figure, frame 229 is shown with different values for the PSNR. These values are close to the PSNR performance presented by the algorithms under evaluation, when there are 50 video application users in the cell, as depicted in Figure 6.19. It is possible to see that the average PSNR presented by RME and RR corresponds to a corrupted frame (Figures 6.20e and 6.20f). As we could not find a frame with the exactly average value for these schedulers (RME = 20 dB, RR = 18 dB), we presented two frames with values that generate those averages values. This corrupted frame is reflection of the metrics used by these two QoS-unaware schedulers.

Figure 6.20d shows the frame for the average PSNR offered by RPQoS. As one can see, despite the fact that RPQoS is a QoS-aware scheduler, when the cell is crowded, the metric of this algorithm could not keep the good performance, and the average PSNR was compromised. We believe that this occurred since the RPQoS metric is based only on the HoL and priority of the users. It does not considers the demand of each user, reported by the BSR.

On the other hand, TSGA presented the best performance (Figure 6.20c). In fact, TSGA provided the same PSNR of the original video transmitted, i.e., after the encoding process (Figure 6.20b). We believe that this is result of the dynamic of our proposed metric, that not only searches for the urgent users, but differentiate these urgent users,

trying to attend their demand. We also believe that these results demonstrate the power of the genetic algorithm search, allied to a smart scheduling metric.



(a) Original. (b) Original, after the enconder (42 dB).

(c) TSGA (42 dB) (d) RP-QoS (30 dB)

(e) RR and RME (24 dB) (f) RR and RME (12 dB)

Figure 6.20: Comparison of the average video quality provided by each scheduling algorithm under evaluation, considering 50 video application users in the cell.

6.3.6 Algorithm Complexity

To perform the analysis of complexity of the algorithms under evaluation, the RR algorithm was used as the baseline. RR is an algorithm that does not depend on any parameter as the number of users (N) or the number of RBs (M) available for resource allocation. Hence, it presents a constant complexity of $O(1)$.

RME and RP algorithms present similar complexity that is dependent on the number of users and RBs in the system. Thereby, they present an overall complexity of $O(N \cdot M)$.

RPQoS has some improvements to assure the QoS requirements. These improvements add complexity to the algorithm, since the algorithm must store and process information about the status of the packets in the UE's buffers. The buffer information is dependent on the number of active users. Therefore, the RPQoS complexity can be approximated to $O(N^2 \cdot M)$.

TSGA is also dependent on N and M. As a QoS-aware algorithm, TSGA has also to store and process information of the UE's buffers. In addition, TSGA must process the GA-based search. The complexity of the GA search can be high if the parameters are not constant and the calculation of the fitness function is complex. In this sense, the proposed algorithm was divided into multiple steps to reduce the complexity of the GA search while keeping its advantages. Limiting the number of users in the Third Step allows the parameters of GA to be constant and the complexity of the search becomes linear.

Another important factor in the complexity of the GA search is the values of the parameters. The number of generations and the size of the population are the parameters that most impact in the convergence time. Hence, we carefully set these parameters, considering a tradeoff between performance and convergence time.

Figure 6.21 shows the average simulation execution time for the algorithms under evaluation. The average execution time was calculated using the same hardware and measuring the interval of time for each scheduler, while it allocates the resources. The average value considers the 50 repetitions of each simulation. From Figure 6.21, one can note that despite in theory TSGA presents the highest complexity for the worst case, in practice, the algorithm presents a competitive execution time. In the evaluation, TSGA is executed less than 1.3 times the execution time required by RR to run. Another important factor for this satisfactory execution time is the action of Step Two. When the network is crowded and the delay of the packets starts to be close to the limit, Step Two directly allocates the resources, avoiding the more time-consuming step of the GA search.

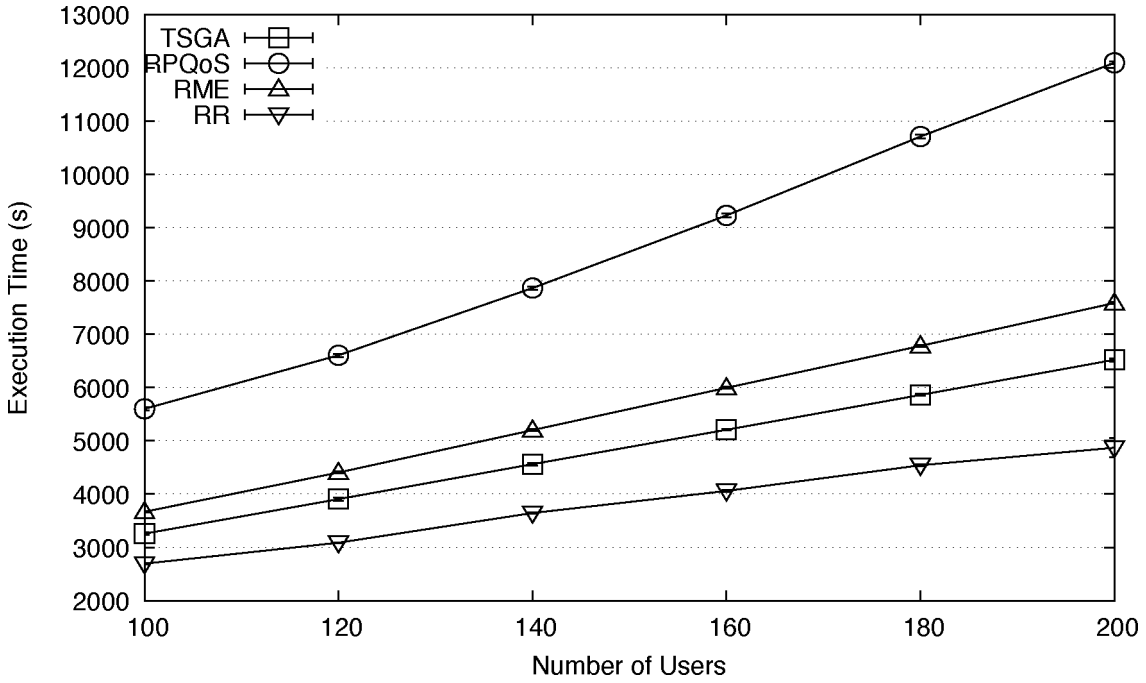


Figure 6.21: Average execution time for simulations.

6.4 Summary

This chapter was dedicated to the performance evaluation of the proposed algorithm. In this sense, ns-3 was chosen as the simulation environment to conduct the evaluation.

Results showed that the Three-Step GA-based algorithm can provide a great performance. TSGA was able to meet the QoS requirements of the GBR applications in the evaluation and caused less pressure in Non-GBR applications, when compared to its competitors.

"I have not failed. I've just found 10,000 ways that won't work."

- Thomas Alva Edison

7

Conclusions

The LTE system is an important wireless mobile technology. It has been developed to meet the always crescent expectations of the users. This network is based on a new all-IP architecture and aims to offer high data rates and low latency. The LTE technology is often branded 4G, but the first release does not fully comply with the IMT-Advanced requirements. This was achieved with the LTE-Advanced release.

To meet these requirements, LTE employs advanced technologies and strategies, such as OFDMA/SC-FDMA, MIMO, Link Adaptation, HARQ, among others. These techniques are used to increase the spectral efficiency and to share the resources of the network in a smart way.

The resource allocation strategies are an important part of the network, since they are responsible to satisfy the different requirements of the applications defined by the QoS classes. In this sense, scheduling algorithms try to meet a system objective, using information of channel's quality, buffer's size, user's energy and QoS requirements, for example. Then, one can note that scheduling algorithms deal with complex optimization problems. The problem is even more complex when we consider the uplink resource allocation, since SC-FDMA requires that the RBs assign to a particular user must be contiguous. Some researchers classify the uplink resource allocation as a NP-hard optimization problem [5].

In this context, it is possible to find several scheduling algorithms for the LTE uplink resource allocation that make use of greedy strategies to allocate the RBs among the users. On the other hand, there are tools created to deal with NP-hard problems. One of the most known is the Genetic Algorithms. To the best of our knowledge, there is no work dealing with the uplink resource allocation using GA as the heuristic of the searching algorithm in complex scenarios with mixed traffics. In this sense, this thesis presented a new Three-Step scheduling algorithm for the LTE uplink resource allocation based on GA.

Despite of the availability of GA-based algorithms for the LTE downlink, it was necessary to create new strategies for some of the genetic operations of the algorithm, since the traditional strategies became obsolete with the incorporation of the contiguity constraint. Then, the design of the proposed algorithm introduced new strategies of initialization, crossover and mutation. Besides the contiguity constraint, the proposed algorithm considered also the impact of the HARQ retransmissions in the allocation.

The Three-Step GA-based algorithm (TSGA) was evaluated in a simulation environment. The ns-3 network simulator was chosen as the environment to perform the evaluation. The RR, RME and RP-QoS algorithms were chosen to participate of the evaluation to verify the relevance of the proposed algorithm. These algorithms were chosen since they are well known algorithms, studied and used as based for several other algorithm proposals. The algorithms were evaluated in scenarios of video chat transmission and compared according to some performance indicators.

Results showed that TSGA was able to meet the QoS requirements of the GBR applications in all scenarios of evaluation. To accomplish that, TSGA had to sacrifice the throughput of the Non-GBR applications. However, the proposed algorithm did not allow Non-GBR applications to enter in state of starvation. Furthermore, TSGA performed the resource allocation in a fairer way than its main competitor, the RP-QoS algorithm.

Considering the algorithm complexity, despite TSGA presented the highest complexity for the worst case scenario, in practice, the proposed algorithm presented a competitive execution time, when compared with the Round Robin algorithm, which is the simplest algorithm in the evaluation. We believe this was possible due to the strategy of resource allocation divided in multiple steps.

As future works, it would be important to conduct evaluations with different values for GA parameters. We also highlight the necessity of an analysis of the Worst Case Execution Time (WCET) to verify the implications of deploying the algorithm in a real network. Finally, it would be also interesting to evaluate the algorithms in other

scenarios with different traffic distribution and verify the behavior of the proposed algorithm in this new context of evaluation.

"If I have seen further than others, it is by standing upon the shoulders of giants."

- Isaac Newton

References

[1] F. Capozzi, G. Piro, L. Grieco, G. Boggia, P. Camarda, Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey, IEEE Communications Surveys & Tutorials 15 (2) (2013) 678–700. doi:10.1109/SURV.2012.060912.00100.

[2] C. Cox, An Introduction to LTE: LTE, LTE-Advanced, SAE and 4G Mobile Communications, Wiley, 2012.

[3] M. E. Aydin, R. Kwan, J. Wu, Multiuser Scheduling on the LTE Downlink with Meta-Heuristic Approaches, Physical Communication 9 (2013) 257–265. doi:10.1016/j.phycom.2012.01.004.

[4] F. D. Calabrese, Scheduling and Link Adaptation for Uplink SC-FDMA Systems A LTE Case Study, Ph.D. thesis, Aalborg University (2009).

[5] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, S. Lu, Proportional Fair Frequency-Domain Packet Scheduling for 3GPP LTE Uplink, in: Proc. of 2009 28th Conference on Computer Communications (INFOCOM), IEEE, 2009, pp. 2611–2615. doi:10.1109/INFCOM.2009.5062197.

[6] N. Abu-Ali, A.-E. M. Taha, M. Salah, H. Hassanein, Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework, IEEE Communications Surveys & Tutorials 16 (3) (2014) 1239–1265. doi:10.1109/SURV.2013.1127.00161.
URL http://ieeexplore.ieee.org/document/6687313/

[7] M. Al-Rawi, R. Jantti, J. Torsner, M. Sagfors, Opportunistic Uplink Scheduling for 3G LTE Systems, in: Proc. of 2007 Innovations in Information Technologies (IIT), IEEE, 2007, pp. 705–709. doi:10.1109/IIT.2007.4430425.

[8] J. Lim, H. Myung, K. Oh, D. Goodman, Proportional Fair Scheduling of Uplink Single-Carrier FDMA Systems, in: Proc. of 2006 17th International Symposium

on Personal, Indoor and Mobile Radio Communications, IEEE, 2006, pp. 1–6. doi:10.1109/PIMRC.2006.254224.

[9] F. D. Calabrese, P. H. Michaelsen, C. Rosa, M. Anas, C. Ubeda Castellanos, C. U. Castellanos, D. L. Villa, K. I. Pedersen, P. E. Mogensen, Search-Tree Based Uplink Channel Aware Packet Scheduling for UTRAN LTE, in: Proc. of 2008 Vehicular Technology Conference, IEEE, 2008, pp. 1949–1953. doi:10.1109/VETECS.2008.441.

[10] F. D. Calabrese, C. Rosa, M. Anas, P. H. Michaelsen, K. I. Pedersen, P. E. Mogensen, Adaptive Transmission Bandwidth Based Packet Scheduling for LTE Uplink, in: Proc. of 2008 Vehicular Technology Conference, IEEE, 2008, pp. 1–5. doi:10.1109/VETECF.2008.316.

[11] L. A. M. Ruiz de Temino, G. Berardinelli, S. Frattasi, P. Mogensen, Channel-Aware Scheduling Algorithms for SC-FDMA in LTE Uplink, in: Proc. of 2008 19th International Symposium on Personal, Indoor and Mobile Radio Communications, IEEE, 2008, pp. 1–6. doi:10.1109/PIMRC.2008.4699645.

[12] H. Safa, K. Tohme, LTE Uplink Scheduling Algorithms: Performance and Challenges, in: Proc. of 2012 19th International Conference on Telecommunications (ICT), no. Ict, IEEE, 2012, pp. 1–6. doi:10.1109/ICTEL.2012.6221230.

[13] F. Z. Kaddour, E. Vivier, L. Mroueh, M. Pischella, P. Martins, Green Opportunistic and Efficient Resource Block Allocation Algorithm for LTE Uplink Networks, IEEE Transactions on Vehicular Technology 64 (10) (2015) 4537–4550. doi:10.1109/TVT.2014.2365960.

[14] M. Salah, N. A. Ali, A.-E. Taha, H. Hassanein, Evaluating Uplink Schedulers in LTE in Mixed Traffic Environments, in: Proc. of 2011 International Conference on Communications (ICC), IEEE, 2011, pp. 1–5. doi:10.1109/icc.2011.5962629.

[15] H. Safa, W. El-Hajj, K. Tohme, A QoS-Aware Uplink Scheduling Paradigm for LTE Networks, in: Proc. of 2013 27th International Conference on Advanced Information Networking and Applications (AINA), IEEE, 2013, pp. 1097–1104. doi:10.1109/AINA.2013.38.

[16] F. Liu, Y.-a. Liu, Improved Scheduling Algorithms for Uplink Single Carrier FDMA System, Journal of Information & Computational Science 11 (2012) 3211–3219.

[17] F. Liu, X. She, L. Chen, H. Otsuka, Improved Recursive Maximum Expansion Scheduling Algorithms for Uplink Single Carrier FDMA System, in: Proc. of 2010

71st Vehicular Technology Conference, IEEE, 2010, pp. 1–5. doi:10.1109/VETECS.2010.5493985.

[18] M. Mitchell, An Introduction to Genetic Algorithms, 5th Edition, MIT Press, 1999.

[19] X. Cheng, P. Mohapatra, Quality-Optimized Downlink Scheduling for Video Streaming Applications in LTE Networks, in: Proc. of 2012 Global Communications Conference (GLOBECOM), IEEE, 2012, pp. 1914–1919. doi:10.1109/GLOCOM.2012.6503395.

[20] N. Sharma, A. S. Madhukumar, Genetic Algorithm Aided Proportional Fair Resource Allocation in Multicast OFDM Systems, IEEE Transactions on Broadcasting 61 (1) (2015) 16–29. doi:10.1109/TBC.2015.2389692.

[21] F. Sun, M. You, J. Liu, Z. Shi, P. Wen, J. Liu, Genetic Algorithm Based Multiuser Scheduling for Single- and Multi-Cell Systems with Successive Interference Cancellation, in: Proc. of 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, IEEE, 2010, pp. 1230–1235. doi:10.1109/PIMRC.2010.5672038.

[22] Chengcheng Yang, X. Xu, Jiang Han, Waheed ur Rehman, Xiaofeng Tao, GA Based Optimal Resource Allocation and User Matching in Device to Device Underlaying Network, in: 2014 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), IEEE, 2014, pp. 242–247. doi:10.1109/WCNCW.2014.6934893.

[23] M. Kalil, J. Samarabandu, A. Shami, A. Al-Dweik, Performance Evaluation of Genetic Algorithms for Resource Scheduling in LTE uplink, in: 2014 International Wireless Communications and Mobile Computing Conference (IWCMC), IEEE, 2014, pp. 948–952. doi:10.1109/IWCMC.2014.6906483.

[24] Saulo da Mata Homepage (2017).
URL http://www.saulodamata.com

[25] S. Sesia, I. Toufik, M. Baker, LTE - the UMTS Long Term Evolution: From Theory to Practice, 2nd Edition, Wiley, 2011.

[26] T. Ali-Yahiya, Understanding LTE and Its Performance, Springer, 2011.

[27] E. Dahlman, S. Parkvall, J. Skold, 4G, LTE-Advanced Pro and The Road to 5G, 3rd Edition, Academic Press, 2016.

[28] ITU-R, Recommendation M.687-2 International Mobile Telecommunications-2000 (IMT-2000) (1997).

[29] Ericsson, Ericsson Mobility Report, Tech. Rep. November, Ericsson (2016).

[30] ITU-R, IMT Vision – Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond, M Series, Recommendation ITU-R M.2083-0 (2015).

[31] H. Holma, A. Toskala, LTE for UMTS: Evolution to LTE-Advanced, Wiley, 2011.

[32] M. Rumney, White Paper Agilent Technologies: De-mystifying Single Carrier FDMA The New LTE Uplink (2008).

[33] 3GPP, 3GPP TS 23.203 V12.0.0 Policy and charging control architecture (2013).

[34] M. Salah, Comparative Performance Study of LTE Uplink Schedulers, Master Thesis, Queen's University (2011).

[35] R. Linden, Algoritmos Genéticos, 3rd Edition, Ciência Moderna, 2012.

[36] 3GPP, TS 36.213 V11.0.0 Physical layer procedures (2012).

[37] T. Issariyakul, E. Hossain, Introduction to Network Simulator NS2, 1st Edition, Springer, 2009.

[38] 3GPP, 3GPP TR 36814 V9.0.0 Further Advancements for E-UTRA Physical Layer Aspects (Release 9) (2010).

[39] Network Simulator 3 Homepage (2017).
URL `http://www.nsnam.org`

[40] N. S. . Consortium, ns-3 Manual (2014).

[41] LENA Module (2017).
URL `http://lena.cttc.es/manual/lte.html`

[42] W. Navidi, T. Camp, Stationary Distributions for the Random Waypoint Mobility Model, IEEE Transactions on Mobile Computing 3 (1) (2004) 99–108.

[43] W. Navidi, T. Camp, N. Bauer, Improving the Accuracy of Random Waypoint Simulations through Steady-State Initialization, in: Proc. of the 15th International Conference on Modeling and Simulation, 2004, pp. 319–326.

[44] B. Bojovic, N. Baldo, A New Channel and QoS Aware Scheduler to Enhance the Capacity of Voice Over LTE Systems, in: 2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14), IEEE, 2014, pp. 1–6. `doi:10.1109/SSD.2014.6808890`.

[45] EvalVid Framework (2017).
URL `http://www.tkn.tu-berlin.de/menue/research/evalvid`

[46] NS-3 EvalVid Module (2017).
URL `https://github.com/gercom/evalvid-ns3`

[47] R. Jain, D. Chiu, W. Hawe, A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems (1984).
URL `http://arxiv.org/abs/cs/9809099`