



HAL
open science

Resources allocation in high mobility scenarios of LTE networks

Jean-Thierry Stephen Avocanh

► **To cite this version:**

Jean-Thierry Stephen Avocanh. Resources allocation in high mobility scenarios of LTE networks. Networking and Internet Architecture [cs.NI]. Université Sorbonne Paris Cité, 2015. English. NNT : 2015USPCD052 . tel-01661372

HAL Id: tel-01661372

<https://theses.hal.science/tel-01661372>

Submitted on 11 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Number assigned by the library

--	--	--	--	--	--	--	--	--	--

RESOURCES ALLOCATION IN HIGH MOBILITY SCENARIOS OF LTE NETWORKS

Jean Thierry Stephen AVOCANH



A thesis submitted for the degree of Doctor of Philosophy

Subject: Networks and Telecommunications

Defended and presented on **October 16th, 2015**

at Laboratoire de Traitement et Transport de l'Information (L2TI)

École doctorale Galilée

Université Paris 13, Sorbonne Paris Cité



Institut Galilée

Members of the jury :



Pr. Jalel BEN-OTHTMAN L2TI, Supervisor

Dr. Marwen ABDENNEBI L2TI, Co-Supervisor

Pr. Samir TOHME PRISM, Reviewer

Pr. Hacène FOUCHAL CReSTIC, Reviewer

Pr. Lynda MOKDAD LACL, Examiner

Pr. Karim DJOUANI LISSI, Examiner

Pr. Boubaker DAACHI MESR - UPL, Examiner



Villetaneuse, France, 2015

N° attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

ALLOCATION DE RESSOURCES RADIO DANS LES RESEAUX LTE A FORTE MOBILITE

Jean Thierry Stephen AVOCANH



Thèse pour l'obtention du grade de Docteur de l'Université Paris 13

Spécialité: Réseaux et Télécommunications

Présentée et soutenue publiquement le **16 Octobre 2015**

au Laboratoire de Traitement et Transport de l'Information (L2TI)

École doctorale Galilée

Université Paris 13, Sorbonne Paris Cité



Institut Galilée

Devant le jury composé de :



Pr. Jalel BEN-OTHMAN L2TI, Directeur de these

Dr. Marwen ABDENNEBI L2TI, Co-Encadrant de these

Pr. Samir TOHME PRISM, Rapporteur

Pr. Hacène FOUCHAL CReSTIC, Rapporteur

Pr. Lynda MOKDAD LACL, Examineur

Pr. Karim DJOUANI LISSI, Examineur

Pr. Boubaker DAACHI MESR - UPL, Examineur



Villetaneuse, France, 2015

Declaration of Authorship

I, Jean Thierry Stephen AVOCANH, declare that this thesis titled, 'Resources allocation in high mobility scenarios of LTE networks' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed:

Date:

Résumé

Cette étude porte sur l'allocation de ressources radio dans les réseaux LTE à forte mobilité. En particulier, il s'agit de concevoir des stratégies d'allocation de ressources capables d'améliorer la qualité de service des flux multimédia dans un contexte de forte mobilité des terminaux. Pour atteindre ces objectifs, l'étude a été menée en deux étapes. Dans un premier temps les travaux se sont déroulés dans un contexte où l'aspect *forte mobilité* n'a pas été pris en compte. Cela a permis de bien maîtriser tous les aspects liés à l'allocation de ressources dans le LTE tout en proposant de nouvelles méthodes meilleures que celles existantes. Une fois cette tâche accomplie, l'aspect *forte mobilité* a été ajouté au problème et des stratégies adaptées à ce contexte ont été proposées. Néanmoins, dû aux différences entre les liens montants et descendants, l'étude a été menée dans les deux directions.

Comme première contribution, nous avons conçu deux stratégies pour améliorer l'allocation de ressources sur la liaison descendante dans un contexte où la forte mobilité n'a pas été prise en compte. La première méthode est un mécanisme qui améliore cette allocation en particulier dans les scénarios *d'overbooking* en faisant un compromis entre l'équité, le débit global du système et les exigences de qualité de service des applications. La seconde stratégie permet non seulement de satisfaire aux contraintes de délais mais également de garantir un très faible taux de perte de paquets aux services de type multimédias. Les performances des systèmes proposés ont été évaluées par des simulations en les comparant à d'autres mécanismes dans la littérature. Les analyses ont démontré leur efficacité et révélé qu'elles obtenaient les meilleures performances.

Notre deuxième contribution a permis d'améliorer l'allocation de ressources toujours dans un contexte de non prise en compte de la forte mobilité, mais cette fois ci sur le lien montant et pour des flux de type vidéo téléphonie. Nous avons conçu un nouveau protocole qui réduit de façon considérable les retards causés par l'allocation dynamique des ressources. L'idée consiste à allouer des ressources à ces trafics en utilisant une stratégie semi-persistante associée à un processus de pré-allocation. Les performances de notre méthode ont été évaluées par simulations et les résultats ont montré qu'elle fournissait le meilleur support en qualité de service.

La dernière partie de nos travaux s'est intéressée au problème d'allocation de ressources dans les scénarios de fortes mobilités des terminaux. Dans cette partie, nous avons élaboré deux stratégies efficaces convenant aux scénarios véhiculaires. La première méthode est une technique permettant de maintenir le niveau de qualité de service nécessaire pour le bon fonctionnement des applications vidéo des utilisateurs ayant les vitesses les plus élevées. Elle consiste à déterminer en fonction des différentes vitesses des utilisateurs, le taux minimum de rapports CQI à envoyer à la station de base. Quant à la seconde stratégie, c'est un procédé d'ordonnancement opportuniste qui améliore la qualité de service des applications vidéo des utilisateurs ayant les vitesses les plus élevées. Avec cette stratégie, ces utilisateurs obtiennent une plus grande priorité et acquièrent ainsi beaucoup plus de ressources.

Abstract

Our thesis focuses on issues related to resources allocation in LTE Networks. In particular the purpose of this study is to design efficient scheduling algorithms to improve the QoS of real time flows in a context of high mobility of the users. To reach this goal, the study has been carried out in two steps. At first, in order to have an expert knowledge of the key facets of LTE scheduling, we conducted the study in a context where the high mobility aspect of the node was not taken into account. This helped not only to critically analyze the literature but also to propose new schemes to improve QoS of real time applications. After that, the high mobility parameter has been added and innovative methods dealing with this context have been designed. Nevertheless due to the existing differences between the downlink and the uplink, the issue was tackled in each of the aforementioned directions.

We firstly addressed the problem of improving the scheduling of downlink communications in a context where the high mobility was not taken into account. Two major methods have been designed for this purpose. The first one is an innovative scheme which improves resources assignment in overbooking scenarios by doing a trade-off between fairness, overall system throughput and QoS requirements. The second one is an enhanced scheduling scheme which provides strict delay bounds and guarantees very low packet loss rate to multimedia flows. The performance of the proposed schemes have been evaluated by simulations and compared to other schemes in the literature. The analyses demonstrated their effectiveness and showed that they outperformed the existing ones.

The second contribution concerned the problem of improving the scheduling of uplink communications in a context where the high mobility was not taken into account. We designed a novel scheduling protocol which improves resources allocation for videotelephony flows and reduces the delay caused by dynamic scheduling. It consists in scheduling such traffics using a semi-persistent strategy associated with a provisioning process. The performance of our proposed method have been evaluated by simulations and results demonstrated its effectiveness by showing that it improved videotelephony flows performance and provided the best QoS support compared to the dynamic scheduling.

The last contribution addressed the problem of resources allocation in high mobility scenarios. In this part, the high mobility aspect was taken into account for designing suitable schemes for vehicular scenarios. We proposed in this way two efficient strategies. The first one is a technique which maintains the required level of QoS for supporting video users at high velocities. It consists in identifying depending on the UEs velocity, the minimum CQI reports rate in order to maintain the required QoS. The second proposed strategy is an opportunistic method which improves the performance of high speed video users. With this strategy, more priority are given to the UEs having the highest velocity. Simulations results demonstrated its effectiveness and showed that it improved the QoS support of video users having the highest velocity.

Acknowledgements

I could not start writing this present manuscript without as a preliminary addressing my sincere thanks to all persons who were of an appreciable contribution to the success of this thesis.

I would like to make a point of my gratitude to my supervisors Pr. Jalel Ben-Othman and Dr. Marwen Abdennebi, for their guidance and understanding;

I would like also to thank Pr. Samir Tohmé and Pr. Hacène Fouchal for kindly accepting to review my work and being part of the jury, especially under such short notice. Their valuable feedbacks helped to improve the dissertation to its current state.

Finally, I am really thankful to all the members of the L2TI lab and particularly to my colleagues for their support, help, advices and friendship.

That all of these people as all of those that the names do not appear within this document and who contributed to the success of this thesis find here the expression of my gratitude.

Contents

Declaration of Authorship	ii
Resume	iv
Abstract	vi
Acknowledgements	viii
Contents	x
List of Figures	xiii
List of Tables	xvi
Abbreviations	xix
1 Introduction	1
1.1 Context of the thesis	1
1.2 Study Organization	6
1.3 Document Layout	7
2 State-of-the-art of LTE Networks	9
2.1 Introduction	9
2.2 System architecture	10
2.2.1 The Core Network	12
2.2.2 The Radio Access Network	13
2.3 QoS Management	14
2.4 Radio protocols	16
2.4.1 Protocol stack layers	17
2.4.1.1 Radio Resource Control (RRC)	17
2.4.1.2 Packet Data Convergence Protocol (PDCP)	18
2.4.1.3 Radio Link Control Layer (RLC)	18
2.4.1.4 Medium Access Control (MAC)	19
2.4.1.5 Physical Layer	20

2.4.2	Communication channels	22
2.4.2.1	Logical channels	22
2.4.2.2	Transport channels	23
2.4.2.3	Physical channels	24
2.5	Radio resources management procedures	26
2.5.1	CQI	26
2.5.2	HARQ	26
2.5.3	AMC and TPC	27
2.6	Scheduling	27
2.6.1	Scheduling process	29
2.6.1.1	Scheduling process in the downlink	29
2.6.1.2	Scheduling process in the uplink	30
2.6.2	Key design aspects	31
2.6.3	Scheduling strategies classification	32
2.6.3.1	Channel-unaware strategies	32
2.6.3.2	Channel-aware/QoS-unaware strategies	34
2.6.3.3	Channel-aware/QoS-aware strategies	36
2.6.3.4	Persistent and semi-persistent strategies	37
2.7	Conclusion	38
3	Resources allocation in the Downlink direction	41
3.1	Introduction	41
3.2	A new two-level scheduling algorithm to face overbooking scenarios	43
3.2.1	Context	43
3.2.2	Literature review	43
3.2.3	Key design aspects	44
3.2.3.1	Metric description	45
3.2.3.2	Algorithm	46
3.2.3.3	Model description	46
3.2.4	Performance analysis	54
3.3	An enhanced scheduling scheme for multimedia services	59
3.3.1	Context	59
3.3.2	Literature review	59
3.3.3	Key design aspects	61
3.3.3.1	The highest level of the scheduler	62
3.3.3.2	The lowest level of the scheduler	62
3.3.4	Performance analysis	64
3.3.4.1	Traffic model	66
3.3.4.2	Results and analysis	66
3.4	Conclusion	70
4	Resources allocation in the Uplink direction	72
4.1	Introduction	72
4.2	A Semi-Persistent scheme for videotelephony traffics	73
4.2.1	Context	73
4.2.2	Uplink transmissions of videotelephony traffics	74
4.2.3	Literature review	78

4.2.4	Key design aspects	79
4.2.4.1	Protocol	79
4.2.4.2	Prediction model	82
4.2.4.3	Scheduling algorithm	89
4.2.5	Performance analysis	89
4.2.5.1	Traffic model	90
4.2.5.2	Results and analyses	92
4.3	Conclusion	94
5	Resources allocation in high mobility scenarios	97
5.1	Context	97
5.2	Vehicular scenarios overview	98
5.3	Effects involved with the high velocity scenarios	100
5.4	Impact of the effects involved with the high velocity scenarios	101
5.5	Proposed methods	105
5.5.1	CQI rate compromise	106
5.5.2	Opportunistic scheduling scheme	108
5.5.2.1	Context and key design aspects	108
5.5.2.2	Performance analysis	109
5.6	Conclusion	112
6	Conclusion & Future directions	115
6.1	Conclusion	115
6.2	Future directions and challenges	118
A	LTE-Sim description	122
B	OTDOA positioning description	124
	Bibliography	126

List of Figures

1.1	LTE subscribers growth forecast from 2012 to 2018	3
1.2	Generic view of a downlink resources scheduler	4
1.3	Difference between OFDMA and SC-FDMA	6
2.1	The Service Architecture Evolution (SAE) in LTE network	12
2.2	LTE Radio protocol stacks	17
2.3	LTE resource structure	21
2.4	Mapping between the different types of channels	25
2.5	Mapping of control information to physical channels	25
2.6	General model of a downlink packet scheduler	30
2.7	Interactions between eNB and UE in uplink scheduling	31
3.1	Flowchart of the proposed algorithm	47
3.2	MCS distribution in the System	48
3.3	The Markov Chain	50
3.4	Blocking Probabilities of traffic classes	53
3.5	Blocking Probabilities of the different configurations	54
3.6	Best user's number of RBs	56
3.7	Best user's throughput	57
3.8	Overall cell throughput	57
3.9	System fairness index	58
3.10	The enhanced two-level scheduler	61
3.11	Example of scheduling process of two real time flows	64
3.12	PLR of VoIP flows under several traffic loads	67
3.13	PLR of Video flows under several traffic loads	67
3.14	CDF of Video packet delays with 40 Users	68
3.15	Average delay of Video flows under several traffic loads	68
3.16	Aggregate throughput of Video flows under several traffic loads	69
3.17	Aggregate throughput of Best effort flows under several traffic loads	69
4.1	Uplink transmissions of videotelephony traffics with dynamic scheduling	75
4.2	Uplink transmissions of videotelephony traffics with SPS-P	80
4.3	Prediction errors in terms of under allocation	87
4.4	Prediction errors in terms of over allocation	88
4.5	Predictions running time	88
4.6	Non-Segmentation based allocation method	90
4.7	PLR of videotelephony users	92
4.8	CDF of video packet delays with 60 users	93

4.9	PLR of VoIP users	93
4.10	Aggregate throughput of Best Effort users	94
5.1	Generic view of vehicular scenarios in LTE network	99
5.2	Example of FCD uploading via LTE network	100
5.3	PLR of video flows under several traffic loads for scenario 1	103
5.4	PLR of video flows under several traffic loads for scenario 2	104
5.5	CDF of video packet delays with 60 users for scenario 1	104
5.6	Average throughput of video flows for scenario 2	105
5.7	Packet loss ratio of video flows under various CQI reports interval	107
5.8	PLR of video flows of 90 kmph users under several traffic loads	111
5.9	CDF of video packet delays of 90 kmph users	111
5.10	Average delay of video flows of 90 km/h users under several traffic loads	112
A.1	Class diagram of LTE-Sim	123
B.1	Illustration of OTDOA in LTE	125

List of Tables

2.1	Main LTE performance targets	11
2.2	Standardized QoS Class Identifiers for LTE	16
2.3	LTE system bandwidth configuration	22
2.4	Logical channels used in LTE	22
2.5	Transport channels used in LTE	23
2.6	Physical channels used in LTE	24
3.1	Summary of the flows distribution	55
3.2	Simulation parameters	65
4.1	Magi characteristics	87
4.2	Simulation parameters	91
5.1	Main simulation parameters	102
5.2	Simulation cases	102
5.3	Simulation parameters	110

Abbreviations

3GPP	3rd Generation Partnership Project
A-GNSS	Assisted Global Navigation Satellite System
AMC	Adaptive Modulation and Coding
AMPS	Analogue Mobile Phone System
ARP	Allocation and Retention Priority
ARQ	Automatic Repeat reQuest
BSR	Buffer Status Reports
CA	Carrier Aggregation
CDF	Cumulative Distribution Function
CN	Core Network
CODEC	COder-DECoder
COMP	COordinated Multi-Point
CQI	Channel Quality Indicator
CRI	CQI Report Interval
CTMC	Continuous Time Markov Chain
DCI	Downlink Control Information
DS	Dynamic Scheduling
EDGE	Enhanced Data rates for GSM Evolution
E-FLS	Enhanced Frame Level Scheduler
eNB	evolved Node B
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
EPC	Evolved Packet Core
E-SMLC	Evolved Serving Mobile Location Centre
FCD	Floating Car Data
FDPS	Frequency Domain Packet Scheduler

FPS	F rames P er S econd
GBAR	G amma B eta A uto R egressive
GBR	G uaranteed B it R ate
GPRS	G eneral P acket R adio S ervices
GSA	G lobal M obile S uppliers A ssociation
GSM	G lobal S ystem for M obile communications
HARQ	H ybrid A utomatic R etransmission r e Q uest
HSPA	H igh S peed P acket A ccess
J-TACS	J apanese T otal A ccess C ommunication S ystem
LTE	L ong T erm E volution
MAC	M edium A ccess C ontrol
MBSFN	M ulticast B roadcast S ingle- F requency N etwork
MCS	M odulation and C oding S cheme
MIMO	M ultiple I nput M ultiple O utput
MME	M obility M anagement E ntity
NGBR	N on- G uaranteed B it R ate
NRT	N on R eal T ime
OFDMA	O rthogonal F requency D ivision M ultiple A ccess
OTDOA	O bserved T ime D ifference of A rrival
OWD	O ne W ay D elay
PAPR	P eak A verage P ower R atio
PDCCH	P hysical D ownlink C ontrol C hannel
PDCP	P acket D ata C onvergence P rotocol
PDSCH	P hysical D ownlink S hared C hannel
PF	P roportional F air
PGW	P acket data network G ate W ay
PLR	P acket L oss R ate
PRB	P hysical R esource B lock
PUCCH	P hysical U plink C ontrol C hannel
PUSCH	P hysical U plink S hared C hannel
QCI	Q oS C lass I dentifier
QoS	Q uality of S ervice
RAN	R adio A ccess N etwork

RAT	R adio A ccess T echnologie
RB	R esource B lock
RBF	R adial B asis F unction
RLC	R adio L ink C ontrol
RN	R elay N ode
ROHC	R Obust H eaders C ompression
RR	R ound R obin
RRC	R adio R esource C ontrol
RRM	R adio R esource M anagement
RT	R eal T ime
SAE	S ervice A rchitecture E volution
SC-FDMA	S ingle C arrier F requency D ivision M ultiple A ccess
SGW	S erving G ate W ay
SI	S ystem I nformation
SINR	S ignal to I nterference plus N oise R atio
SISO	S ingle I nput S ingle O utput
SPS-P	S emi P ersistent S cheduling with P rovisioning
SR	S cheduling R equest
SRS	S ounding R eference S ignal
SVMs	S upport V ector M achines
TACS	T otal A ccess C ommunication S ystem
TDPS	T ime D omain P acket S cheduler
TTI	T ransmission T ime I nterval
UE	U ser E quipment
UCI	U plink C ontrol I nformation
UMTS	U niversal M obile T elecommunication S ystem
VAD	V oice A ctivity D etector
VBR	V ariable B it R ate

My thoughts go to your place, you who always supported and carried me in your heart: To you my parents, my relatives, my friends and my fiancée. Thanks ever so much for all.

“ It will require doing sacrifices and make you work harder. However, you will never regret and see the benefits at the end.”

Dr. Jacques Traoré
Advices to a young PhD candidate

Chapter 1

Introduction

1.1 Context of the thesis

For years, mobile network operators have engaged a frantic race to reach the best performance throughout the highest data rates, the best coverage and smallest equipments. In this way appeared technologies that have never ceased to be improved in order to face all these challenges.

This evolution was obviously connected with the users requirement which was initially to send urgent messages. Then this has evolved from the simple need to talk with their relatives or contacts to more sophisticated services such as data exchange for which the quantity and quality have never ceased to change.

Faced with these new applications and needs, operators used several successive generations of mobile networks. The first and second generation were dedicated to voice telephony. The first generation (1G) which arrived in the 1980s used analogue technology and comprised a number of independently developed systems worldwide such as AMPS (Analogue Mobile Phone System, used in America), TACS (Total Access Communication System, used in parts of Europe) and J-TACS (Japanese Total Access Communication System, used in Japan). The second generation (2G) system known as GSM (Global System for Mobile communications) was based on digital technology and offered the possibility of global roaming. This expanded over time to take into account data communications, first by circuit-switched transport, then by packet data transport via commercial deployments of GPRS (General Packet Radio Services) and EDGE (Enhanced Data rates for GSM Evolution) in 2000 and 2002 respectively. This helped afterwards to introduce and deploy the third generation (3G) system known as UMTS (Universal Mobile Telecommunication System). UMTS was designed to support multimedia services by providing higher data rates compared to previous technologies. For instance,

theoretical peak data rate for UMTS was about 2 Mbps and those for EDGE was about 472 kbps.

From that moment on, the number of the mobile subscribers has increased hugely. Indeed, the first billion of subscribers was exceeded in 2002, the second billion in 2005, the third billion in 2007 and the fourth billion by the end of 2008. It can be explained by the fact that the mobile has become the preferred way for voice communication due to low cost mobile phones and efficient network coverage and capacity. At the same time we have witnessed an *explosion* of data usage in those mobile networks with a massive uptake of smartphones by subscribers. This was due to the emergence of new applications such as VoIP, Videotelephony, Multimedia Online Gaming, mobile TV and streaming contents. This growing demand for networks services with constraints on delays and bandwidth requirements, posed new challenges in the design of the future generation of mobile networks.

High Speed Packet Access (HSPA) system which is an evolution of UMTS was deployed in 2007 to face this demand but along the way, we have noticed that the end users required higher data rates. For instance, the average data consumption exceeds hundreds of Megabytes per subscriber per month [1]. It was clear that a new and advanced mobile network of higher data capacity with low cost of data delivery was requested.

Thanks to the third Generation Partnership Project (3GPP)¹ which introduced in 2004 the Long Term Evolution (LTE) specifications as an answer to this need [2], aiming at ambitious performance goals and defining new packet-optimized and all-IP architectures for the Radio Access Network (RAN) and the Core Network (CN). It took five years from setting the system targets to the first commercial deployment in Oslo and Stockholm in december 2009². Since the introduction of LTE, its subscribers have never stop growing in the world. At the moment, according to data released by the Global Mobile Suppliers Association (GSA)³ in the latest update of the *Evolution to LTE report*, more than 393 LTE networks have been launched by the cellular operators worldwide in 138 countries and more than 1 billion subscribers are forecast by 2018 as illustrated in Figure 1.1. Starting from this *premise*, it is clear that the *optimization* of all LTE aspects is a topic worth of investigation. Therefore, both research and industrial communities are making a considerable effort on the study of LTE systems, proposing new and innovative solutions in order to analyze and improve their performance.

LTE access network based on Orthogonal Frequency Division Multiple Access (OFDMA) in the downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) in

¹<http://www.3gpp.org>

²TeliaSonera was the first operator to open an available LTE service in the world

³<http://www.gsacom.com>

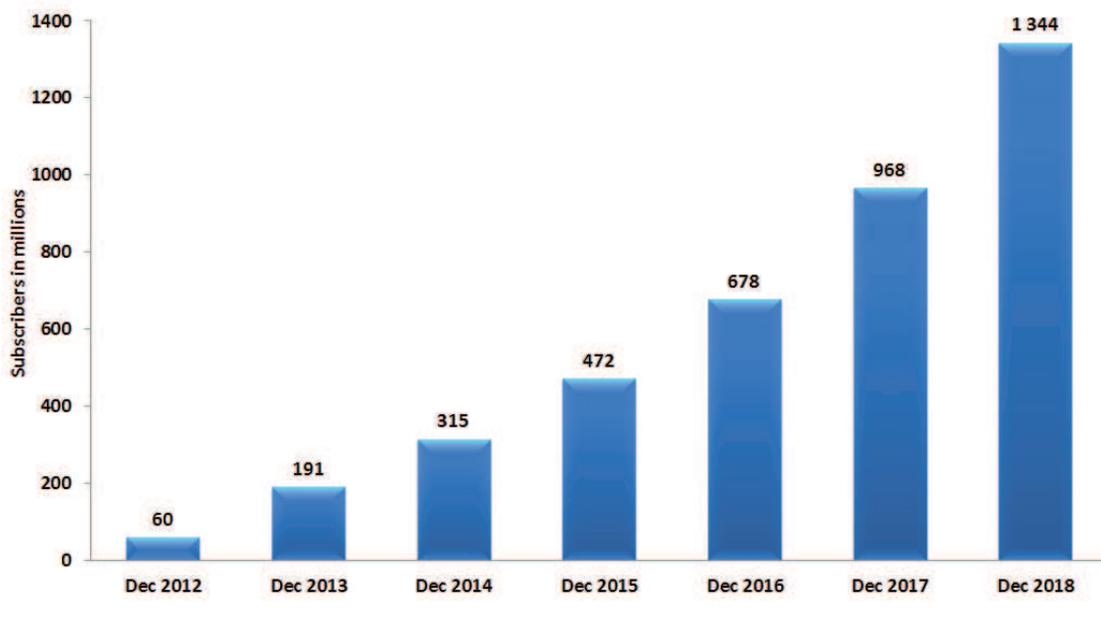


FIGURE 1.1: LTE subscribers growth forecast from 2012 to 2018
Figure from <http://www.statista.com>

the uplink, is expected to provide not only an improved spectral efficiency with respect to previous 3G mobile networks, but also high data rates and low latency even for users in high mobility scenarios [3]. These users can be users in cars or in public transportation having mobile devices, but also vehicles themselves equipped with their own communication interfaces. They represent a significant challenge for networks operators, given the combination of their large data volumes, elevated speed and unique movement patterns.

In order to achieve the required goals, the Radio Resource Management (RRM) block of LTE uses a set of advanced MAC (Medium Access Control) and *physical functions*, like resources sharing, Channel Quality Indicator (CQI) reporting, link adaptation through Adaptive Modulation and Coding (AMC) and Hybrid Automatic Retransmission Request (HARQ) [4]. In this context, the conception of effective resources allocation strategies becomes vital. In fact, more the radio resources will be used efficiently, better the system performance targets will be met and users needs will be satisfied according to specific Quality of Service (QoS) requirements [5].

The *packet scheduler* which works at the radio base station, namely the evolved Node B (eNB), is in charge of the resources allocation functions, as shown in Figure 1.2. Its main task consists in allocating part of spectrum shared among users for both downlink and uplink. In other words, the ultimate aim of its function is typically to fulfil the expectations of as many users of the system as possible, taking into account several parameters such as the radio link quality situation of different users, the QoS requirements, the service priorities and so on. However, the details of the scheduling

process are not standardized as it is largely internal to the eNB, allowing for cutting-edge algorithms to be developed and be optimized for specific scenarios. Considering the major role played by resources allocation in LTE networks and the great interest that high mobility scenarios arouse, a deep study on this attractive topic proves to be necessary.

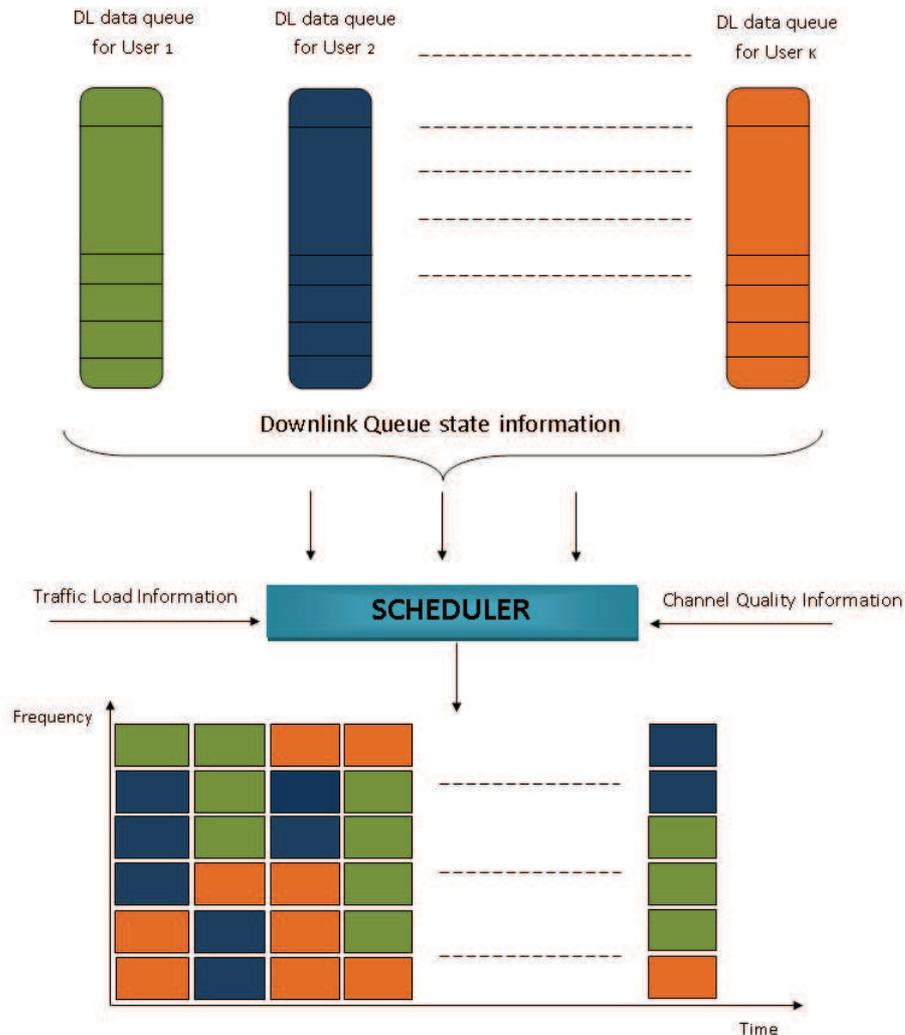


FIGURE 1.2: Generic view of a downlink resources scheduler

Resources allocation in general consists in designing Call Admission Control (CAC) strategies and scheduling algorithms [6]. However, in our study, we essentially focus on issues associated with scheduling algorithms. In particular, the scope of our study is about scheduling in high mobility scenarios of LTE networks. More precisely, the purpose of our work is to design efficient scheduling algorithms to improve the QoS of real time applications in a context where the users have a very high mobility. The challenge of this topic can be found in proposing innovative algorithms in order to fulfill

the QoS requirements of real time multimedia applications, while taking into account both the new concepts introduced by LTE standards such as OFDMA in the downlink and SC-FDMA in the uplink, and the high velocity aspect of the terminals.

The great interest to satisfy the QoS requirements of the multimedia applications is motivated by the fact that they represent the most used applications in telecommunication. Indeed, as mentioned above, the major need of the mobile subscribers falls on the usage of a huge amount of mobile data for video applications such as videotelephony, mobile TV and streaming contents. This is why it is predicted that two-thirds of the world's mobile data traffic will be video by the end of 2015 [7]. However these applications require strict constraints on packet delays and bandwidth requirements for an optimal performance. Therefore, it is very important and necessary to provide the best QoS support to these applications in order to satisfy the subscribers. This is achieved by the conception of appropriate scheduling strategies that improve the performance of these applications.

The interest for high velocity scenarios is due to the fact that users in the network can be vehicles equipped with their own communication interfaces. For this type of nodes, the network enables the use of applications which aims for instance at increasing the passengers safety, at improving also the driving experience or at providing entertainments. The network will also enable the use of new types of applications which will allow the future vehicles, entirely automated, to operate alone. However, the velocity of these particular nodes is high and can reach 193 km/h as mentioned in [8]. As consequences the network topology changes rapidly and system performance can be subject to degradations [9]. Therefore, the network could not be able to provide strong QoS support to the applications of this type of users. In order for the network to provide the best QoS support to this type of users, it is important and necessary to design effective scheduling schemes adapted to high velocity scenarios. These schemes could permit to reduce these performance degradations and thus, improve the QoS support to the vehicles applications.

To reach this goal, the study will be carried out in two steps. At first, in order to have an expert knowledge of the key facets of LTE scheduling, we will conduct the study in a context where the high mobility aspect of the node is not taken into account. This will help not only to reduce the complexity of the work but also to critically analyze the existing works and propose new solutions to improve QoS of real time applications. After that, the high mobility parameter will be subsequently added and also suitable methods dealing with this context will be developed for QoS enhancement of real time applications.

Nevertheless resources allocation has to be done in two direction, namely in downlink and in uplink. Indeed, the designing of efficient resources allocation strategies should not be restricted only to the downlink direction. It is necessary and vital to look into the uplink direction because, the conception of effective scheduling algorithms in this direction is more complicated and represents a great challenge since the proposed methods should not only deal with the constraints related to the scheduling process, but should also be compatible with the constraints imposed by this direction. Indeed, SC-FDMA which is used in uplink requires the allocation to be done exclusively *per users* instead of *per flows* (downlink) and needs also the allocated resources blocks to be contiguous in the frequency domain. Besides, the transmission process requires the eNB to be continuously informed about the amount of the users data to be transmitted, since it is not aware of this amount like in downlink. These constraints oblige scheduling strategies to be modified and redesigned in order to be efficient in the uplink direction. Considering these differences which are illustrated in Figure 1.3, appropriate scheduling schemes will be also designed both at the downlink level and at the uplink level.

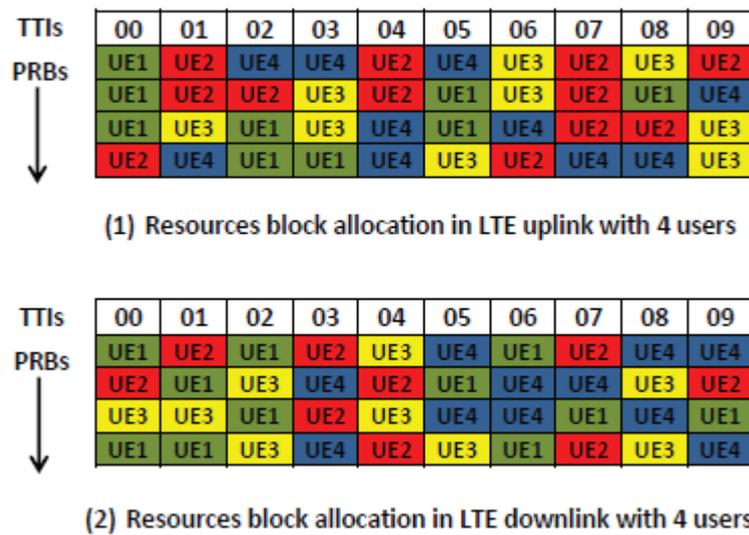


FIGURE 1.3: Difference between OFDMA and SC-FDMA

1.2 Study Organization

Our research falls into the category of scheduling in LTE networks. As presented above, we address two main issues regarding this study, namely scheduling in a context where the high mobility aspect of the users is not taken into account and scheduling in a

vehicular scenario. Due to the existing differences between the downlink and the uplink, the issues are tackled in each of the aforementioned directions.

The work has been conducted using a multidisciplinary approach, including not only Wireless Networking and Computer Networks, but also Programming, Machine Learning and Mathematical Modelling. The main contributions of this thesis are:

- A new two-level scheduling algorithm that optimizes resources assignment in the downlink, in case of overbooking scenarios, by doing a trade-off between spectral efficiency, QoS requirements and fairness.
- An enhanced scheduling mechanism which provides strict delay bounds and guarantees very low packet loss rate to multimedia services in the downlink.
- An innovative semi-persistent scheduling protocol in the uplink direction for videotelephony traffics which improves their performance and reduces the delay caused by dynamic scheduling.
- A new opportunistic scheduling method which improves the performance of high speed video users.

All these contributions will be explained and evaluated in details through this report.

1.3 Document Layout

The first chapter gives an overview of the thesis context, goals and organization. The next chapters are structured as follow:

Chapter 2. State-of-the-art of LTE networks: An overview of LTE system is provided, focusing on the architectural and physical aspects that are relevant from the point of view of the resources allocation problem. In addition, the key issues regarding the design of a scheduler for LTE as well as a detailed classification of the most known strategies are presented for both downlink and uplink.

Chapter 3. Resources allocation in the downlink direction: This section presents our proposed methods to improve downlink resources allocation in a context where the high mobility aspect of the users is not taken into account. The first one is an algorithm that optimizes resources assignment in case of overbooking scenarios, by doing a trade-off between spectral efficiency, QoS requirements and fairness. The second one is an enhanced mechanism which provides strict delay bounds and guarantees very low packet loss rate to multimedia services.

Chapter 4. Resources allocation in the uplink direction: This section presents our semi-persistent scheduling protocol to face the uplink resources allocation issue of videotelephony traffics in a context where the high mobility aspect of the users is not taken into account.

Chapter 5. Resources allocation in high mobility scenario: This chapter presents our solutions to resolve the resources allocation problem in a high mobility context by proposing new schemes which improve the QoS support to video applications of users at high velocity.

Chapter 6. Conclusion and future directions: The last chapter firstly draws the conclusion with particular attention to the lessons learned and after depicts both new research directions and open design challenges.

Chapter 2

State-of-the-art of LTE Networks

This chapter provides an overview of LTE features, focusing firstly on the architectural and technical aspects that are relevant from the point of view of the resources allocation problem. This will help for a better understanding of the fundamentals of the scheduling challenge. In addition, the key issues regarding the design of a scheduler for LTE as well as a detailed classification of the most known and used strategies for LTE scheduling are presented for both downlink and uplink.

2.1 Introduction

As explained in Chapter 1, the emergence of new applications such as VoIP, Videotelephony, multimedia online gaming, mobile TV, Web 2.0 and streaming contents has led to a growing demand for networks services with constraints on delays and bandwidth requirements. This growing demand posed new challenges in the design of the mobile networks future generation. For this purpose, LTE networks have been designed as an answer to this need with very ambitious requirements and aim at performing better than the previous networks. For instance, LTE has been conceived with the goal of evolving the radio access technology. Indeed, in LTE networks all the services are packet-switched instead of circuit-switched [2].

As minimum requirement, LTE network intends not only to provide a spectral efficiency 2 - 4 times better than the one of the previous HSPA networks, but also to enhance the network coverage in terms of bitrate for cell-edge users [10]. Besides, several new performance targets, with respect to previous mobile networks, have been addressed during the different steps of the standardization phase. These performance targets make LTE networks more efficient than the other technologies. As examples of

these performance targets summarized in Table 2.1, LTE networks must provide higher data rates and the support of very high user mobility. Also the latency must be strongly reduced in order to improve the end user performance. In addition, the terminal power consumption must be minimized to enable more usage of the multimedia applications without recharging the battery.

In order to fulfill these requirements, LTE introduced new concepts, spanning from the use of a new network architecture and QoS management policy to major changes at the level of the radio access network. Regarding the new radio access network, it is based on new technical principles such as the use of new multiple access schemes on the air interface and the use of advanced and simplified radio interface protocols. As for the new policy of the QoS management, it is achieved through the management of the radio bearers.

Among all the new concepts introduced in LTE, the most important remains the new advanced Radio Resource Management (RRM) techniques. Indeed, the RRM function aims at using the radio resources in an efficient way and serving the users according to their configured QoS parameters, by exploiting the wide range of the available adaptation techniques and procedures. Thus, an effective use of the RRM mechanisms will allow to satisfy the users needs and meet the system performance requirements. The RRM function is performed for the most part by the scheduling process which consists mainly in designing cutting-edge algorithms which take into account both the QoS requirements and the physical constraints.

Details about all these aforementioned features are important and necessary to be presented. Therefore in this chapter, we will provide a *State-of-the-art* of LTE network by focusing on these new introduced concepts and features. We will start with the description of the new network architecture and then detail how the QoS is managed in the network. Once this task accomplished, we will come closer to the radio access network and give more information about its new technical principles. We will begin with a complete description of the radio protocols and their specificities and after, the RRM procedures will be addressed. Since the scheduling process is the most important procedure of the RRM function, more attention will be given to it afterwards.

2.2 System architecture

LTE has been designed to support only Packet-Switched (PS) services, instead of the Circuit-Switched (CS) model used in the earlier systems. The objective of the LTE network is not only to provide an IP connection with no breaks or gaps between the User

TABLE 2.1: Main LTE performance targets [11]

Peak Data Rate	<ul style="list-style-type: none"> - DL: 150 Mbps (2x2 MIMO and 20 MHz bandwidth) - UL: 75 Mbps (20 MHz bandwidth)
Spectral Efficiency	2 - 4 times better than HSPA networks
User Plane Latency	Below 5 ms for 5 MHz bandwidth or higher
Mobility	<ul style="list-style-type: none"> - Optimized for low mobility up to 15 km/h - High performance for speed up to 120 km/h - Maintaining connection up to 350 km/h
Scalable Bandwidth	From 1.4 to 20 MHz
Duplexing	FDD, TDD and half-duplex FDD
RRM	<ul style="list-style-type: none"> - Enhanced support for end-to-end QoS - Efficient transmissions and operations of higher layer protocols
Services Support	<ul style="list-style-type: none"> - Efficient support of several services such as web-browsing, file transfer and video streaming - VoIP should be supported with at least a good quality as voice traffic over the UMTS networks

Equipment (UE) and the Packet Data Network (PDN), but also to maintain it during the high mobility of the UE [10]. The LTE network architecture which is also known as the Service Architecture Evolution (SAE), is based on a flat architecture [12]. As depicted in Figure 2.1, it is comprised of the following two main components:

- The Evolved-Universal Terrestrial Radio Access Network (E-UTRAN) which is the Radio Access Network (RAN);
- The Evolved Packet Core (EPC) which is the Core Network (CN);

This section will give details on the functions provided by each of the two main network components.

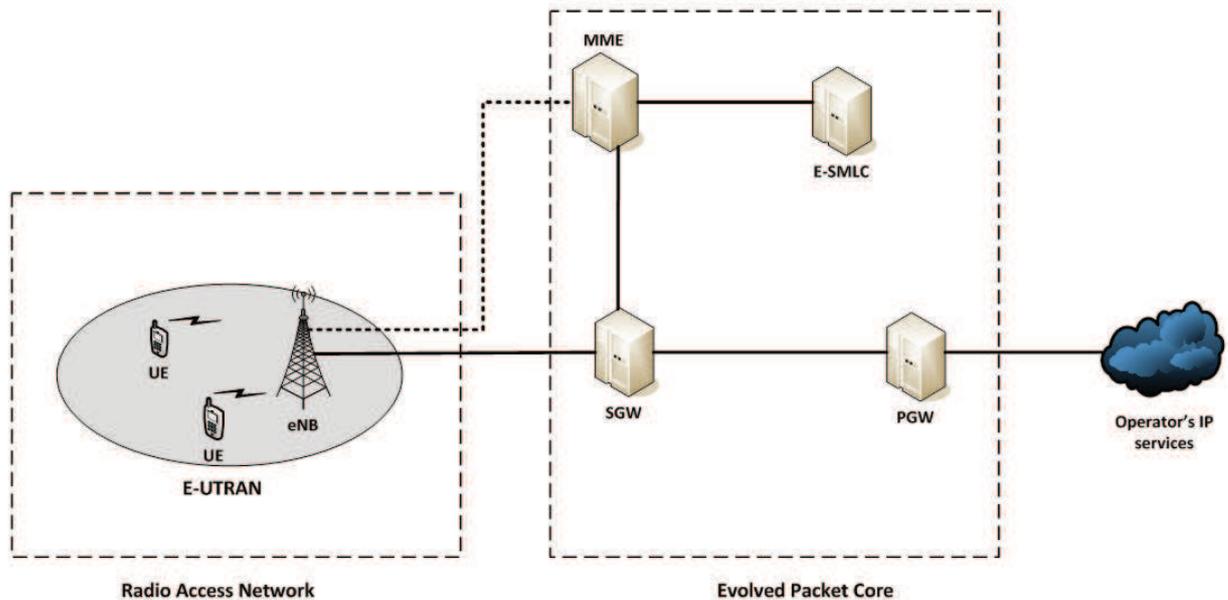


FIGURE 2.1: The Service Architecture Evolution (SAE) in LTE network

2.2.1 The Core Network

The CN is in charge of the overall control of the UE and the setting up of the bearers¹. It is comprised of the following important logical nodes:

- The Packet Data Network Gateway (PGW);
- The Serving GateWay (SGW);
- The Mobility Management Entity (MME);
- The Evolved Serving Mobile Location Centre (E-SMLC).

¹A bearer is a logical packet transmission path

The PGW is in charge of all the IP packet based operations. For instance it allocates IP address to the new UEs connected to the network and carries out the deep inspection of these IP packets. It is also in charge of the QoS enforcement and the flow-based charging. It interconnects the LTE network with the outside world, providing communications among UEs and external PDNs.

The SGW instead, acts as a router and is in charge of forwarding data between the eNB and the PGW. It serves also as a local mobility anchor for inter eNB handovers, as well as the mobility anchor for inter-working with other 3GPP technologies.

The MME processes the signalling operations between the UE and the CN. It is responsible for the UEs mobility, intra-LTE handovers, tracking and paging mechanisms of UEs on connection setting up.

The E-SMLC plays an important role since it manages the overall coordination and scheduling of resources required to find the location of a user connected to E-UTRAN. It also computes the final UE location based on the estimates it receives, and derives the UE speed and the achieved precision.

2.2.2 The Radio Access Network

The E-UTRAN is in charge of handling the radio communications between the UE and the CN. Its main functions are:

Resources management : It encompasses operations such as the radio bearer control, the radio admission control, the radio mobility control and the scheduling of radio resources to UEs in both uplink and downlink.

Header compression : With this function, the RAN aims at efficiently reducing the IP packets headers.

Security : It allows to encrypt all data exchanged over the radio interface.

Positioning : Its role is to help and assist the E-SMLC by providing all the necessary informations for locating the UE.

Connectivity to the CN : It consists of the signalling operations towards the MME and the SGW.

The E-UTRAN comprises of only two types of nodes, namely the UE and the eNB. Due to the fact that there is no centralized controller in LTE, its architecture is said to be *flat*.

Regarding eNBs, they are all inter-connected with each other and connected to the MME and the SGW. Their role has increased significantly in LTE compared to previous mobile networks. Indeed, the eNBs are the only nodes in charge of managing the radio resources and carrying out the control operations.

As for UE, it is the device that the end user uses for communications. It can be a hand-held device such as a smartphone, or it could be embedded to a laptop or a vehicle. The UE contains the Universal Subscriber Identity Module (USIM) which is an application used to identify and authenticate the UE to the network. It is also used for deriving the security keys to protect the radio interface transmissions [1].

In this section we presented the new system architecture with its main components and their different functions. As mentioned since the beginning of this work, LTE has been designed to provide the best QoS support to the UEs, with respect to the previous mobile networks. This particularity is based on the definition and the use of a new QoS management policy. It is important to give more details about this new policy. So the section below presents how the QoS is managed in LTE networks.

2.3 QoS Management

The QoS management in LTE is performed through the management of the radio bearers. A radio bearer is a logical channel established between UE and eNB and associated with a particular QoS. When these bearers are used to carry the control messages they are called Signaling Radio Bearers (SRB) and when they are used for conveying user data, they become Data Radio Bearers (DRB). With the possibility to associate a particular QoS to the radio bearers, QoS differentiation and provision can be achieved through their management.

For an illustration we take the example of a UE running multiple applications with different QoS requirements. For instance, the UE can be engaged in a VoIP session while at the same time browsing a web page or downloading a file. VoIP application requires strict constraints on packets delay than web browsing, whereas file transfer has more constraints on packets loss since it is packets loss sensitive. In order to support multiple QoS requirements, different type of bearers are set up, each being associated with a particular QoS.

When a UE gets connected to the PDN, a *default bearer* is created and given for basic connectivity and exchange of control messages. This bearer remains throughout the entire lifetime of the connection. Another type of bearers called *Dedicated bearers* are created after the UE connection and set up every time a new particular service is

issued. Depending on their QoS requirements, the dedicated bearers can be classified as Guaranteed Bit-Rate (GBR) or Non-Guaranteed Bit-Rate (N-GBR) bearers [13]. We want to point out that the default bearers are always N-GBR bearers.

GBR bearers : These bearers have an associated GBR value for which dedicated network resources are permanently allocated at bearer establishment or modification. If the service requires higher bit rates than the GBR, it can be allowed if resources are available. In this context, a Maximum Bit Rate (MBR) parameter is associated with the GBR bearer and defines the upper limit of the bit rate which can be expected. These types of bearers are used for real time applications such as videotelephony.

Non-GBR bearers : Different from the previous GBR bearers, these bearers have no guaranteed bit rates. In other words, no bandwidth resources are allocated permanently to these bearers. These type of bearers are most of the time used for applications which are not delay sensitive such as web browsing or file transfer.

Each bearer has a QoS profile which includes a set of QoS parameters depending on the application data for which the bearer has been set up. These set of QoS parameters enables QoS differentiation among the flows. The QoS parameters consist of a QoS Class Identifier (QCI) and an Allocation and Retention Priority (ARP).

A QCI is a standardized scalar which is used to access parameters that control the bearer level packet forwarding treatment. In this way, each QCI is characterized by a priority level, a packet delay budget and an acceptable packet loss rate. The value of the QCI of the bearer determines the way it is handled by the eNB. 3GPP has defined during the specifications of LTE several classes of QoS services through QCIs [14]. The set of the standardized QCIs and their characteristics are provided in Table 2.2.

As for the ARP parameter, it is a scalar which contains information about the pre-emption capability and the pre-emption vulnerability of the bearer. It is typically used for call admission control. Indeed, it allows to decide whether a requested bearer establishment or modification can be accepted or needs to be rejected in case of radio resources limitations.

As seen throughout this section, LTE has defined a new policy for managing QoS. QoS differentiation and provisioning is accomplished through the management of bearers. Different types of bearers with different QoS profiles are used for this purpose. Nevertheless, it is important to know the way these bearers are established, configured or even released in the radio access network in order to efficiently support various QoS

TABLE 2.2: Standardized QoS Class Identifiers for LTE

QCI	Resource type	Priority	Packet delay budget [ms]	Packet loss rate	Example of services
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational video (live streaming)
3	GBR	5	300	10^{-6}	Non-conversational video (buffered streaming)
4	GBR	3	50	10^{-3}	Real time gaming
5	Non-GBR	1	100	10^{-6}	IMS signaling
6	Non-GBR	7	100	10^{-3}	Voice, video (live streaming), interactive gaming
7	Non-GBR	6	300	10^{-6}	Video (buffered streaming)
8	Non-GBR	8	300	10^{-6}	TCP based (e.g. WWW, e-mail), chat, FTP, P2P file sharing
9	Non-GBR	9	300	10^{-6}	

class of services. These functions are achieved by the means of the radio protocols which define the way for processing the information flows. The next section will give more informations about the LTE radio protocols with particular attention to the different layers of the protocol stack and the communication channels.

2.4 Radio protocols

As mentioned above, the LTE radio protocols provide the means for setting up, reconfiguring and releasing the radio bearers which are used to transfer applications data or control informations between the UE and the network. The LTE radio protocols are terminated between the eNB and the UE, and the architecture can be separated into *control plane* architecture and *user plane* architecture as depicted in Figure 2.2.

The user plane consists of the Physical layer, the Medium Access Control (MAC) layer, the Radio Link Control (RLC) layer and the Packet Data Convergence Protocol (PDCP) layer. As for the control plane, it is configured from the same protocols as the user plane to the Radio Resource Control (RRC) protocol. In both cases, the information is processed by the different layers before transmission. This section describes the main

functions provided by the radio interface protocols and the way the applications data and control messages are processed and transported by the different layers.

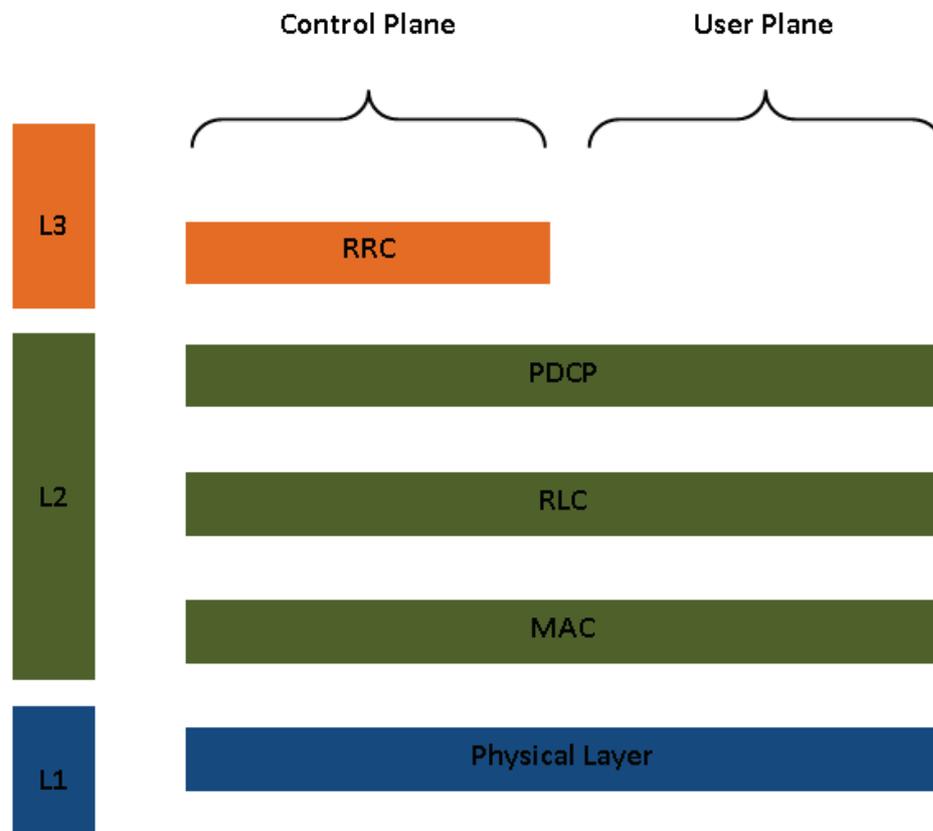


FIGURE 2.2: LTE radio protocol stacks

2.4.1 Protocol stack layers

There are five main radio interface protocols in LTE. We give below the functionalities of each of these radio protocols.

2.4.1.1 Radio Resource Control (RRC)

RRC is responsible for handling layer 3 procedures and supports the transfer of the control messages exchanged between the UE and E-UTRAN [15]. It is also in charge of lower layer configuration. Its main functions include :

- Broadcast of system information;
- Paging;

- Establishment, maintenance and release of an RRC connection between the UE and E-UTRAN;
- Security functions including keys management;
- Establishment, configuration, maintenance and release of the point-to-point radio bearers;
- UE measurements reporting and control;
- Mobility functions (handover);
- UE cell selection and reselection and their control;

2.4.1.2 Packet Data Convergence Protocol (PDCP)

The Packet Data Convergence Protocol (PDCP) processes RRC messages in the control plane and IP packets in the user plane [16]. Its key functionalities are:

Header compression and decompression of IP packets : PDCP layers is responsible for IP header compression and decompression to avoid dispensable overhead in the payload. It uses the Robust Header Compression (ROHC) protocol for this task. Header compression is necessary for applications such as VoIP which have smaller IP packets. Indeed, the large IP header could be a significant source of overhead for small data rates.

Security functions : It is responsible for ciphering and deciphering both the user plane data and the control plane messages, and also for integrity protection and verification of the control plane messages to ensure that messages exchanged between the different nodes come from the right source.

Handover support functions : The PDCP layer is used in two ways during the handover procedures. In the uplink direction, PDCP ensures the retransmission of all the packets which have not been indicated by lower layers to be completed. In the downlink direction, it forwards the non-delivered packets to the target eNB. This is to ensure that no data will be lost during the handover procedures.

2.4.1.3 Radio Link Control Layer (RLC)

The RLC layer is in charge of the concatenation, the segmentation and the reassembly of upper layer packets. It performs these operations to adapt the packets to the size that can actually be supported over the air interface. It is also responsible for errors

detection and correction through the Automatic Repeat reQuest (ARQ) operation [17]. Besides, the RLC layer performs data reordering and duplicate detection. The RLC operates in three modes of data transmission :

Transparent Mode (TM) : TM is the simplest RLC mode. In the TM mode, the RLC only delivers and receives the Protocol Data units (PDUs) on a logical channel without any modifications. In other words, In this mode, RLC does not add or remove any header to the input data and also does not split it into multiple segments. The TM mode is not used for user plane data in LTE and is suitable for services which are not sensitive to delivery order, such as broadcast SI messages and paging messages.

Unacknowledged Mode (UM) : In UM mode, RLC provides a unidirectional data transfer service. In this mode, no ACK or NACK message is required. It is mostly used for data transmission of delay-sensitive Real Time (RT) applications such as VoIP and videotelephony.

Acknowledged Mode (AM) : The AM mode is like the TCP transport mode in the TCP/IP model and provides a bidirectional data transfer service. The most important feature of AM RLC is done through the *retransmissions* procedure. The ARQ operation is achieved to support error-free transmission. Consequently, The AM mode is for the most part used for data transmission of error-sensitive and delay-tolerant Non Real Time (NRT) flows, namely web browsing and file transfer application. In addition, RRC messages use the AM mode for a reliability purpose in the control plane. Indeed, the RLC acknowledgements and retransmissions is used to ensure the reliability of the control messages.

2.4.1.4 Medium Access Control (MAC)

The MAC layer is the lowest sublayer in the Layer 2 architecture of the LTE radio protocol stack and provides the most important procedures for the radio interface [18]. It is responsible for :

Multiplexing/Demultiplexing of data from different radio bearers : It performs the multiplexing of RLC PDUs belonging to one or different radio bearers onto Transport Blocks (TBs) to be delivered to the physical layer. It also performs the reverse operation which consists in the demultiplexing of MAC SDUs from one or different radio bearers from TBs delivered from the the physical layer.

Uplink and downlink scheduling of shared channel resources : The scheduler is located in the eNB and distributes the available radio resources in one cell among the UEs, and among the radio bearers of each UE.

Discontinuous Reception (DRX) : The MAC layer performs the DRX process while setting up an RRC connection for extending the battery lifetime of the UE.

The MAC layer is also responsible for errors correction through Hybrid ARQ (HARQ) operations and reporting of the scheduling informations.

2.4.1.5 Physical Layer

The physical layer is in charge of carrying all informations coming from the MAC layer over the air interface. It also protects data against channel errors using the adaptive modulation and coding (AMC) scheme and handles the modulation and demodulation processes. In addition to these functions, it manages measurement reports from the UE such as Channel Quality Indicator (CQI) for the upper layers.

The physical layer consist of a mixture technologies. Regarding the radio access scheme, it is based on the basic Orthogonal Frequency Division Multiplexing (OFDM) scheme. In particular, SC-FDMA and OFDMA are used in uplink and downlink directions, respectively. They are different from the OFDM scheme since they allow multiple access by assigning sets of subcarriers to each individual UE. The difference between OFDMA and SC-FDMA can be explained by the fact that OFDMA takes advantages of the subcarriers distributed inside the entire spectrum, whereas SC-FDMA uses only adjacent subcarriers. Besides, OFDMA provides high scalability and high robustness against the multipath fading. In contrast, SC-FDMA has a low Peak To Average Power Ratio (PAPR) which allows to increase the power efficiency of UEs and improve the coverage and the cell-edge performance [11].

As for modulation schemes, Quadrature Phase-Shift Keying (QPSK), 16 QAM (Quadrature Amplitude Modulation) and 64 QAM are used both in the downlink and the uplink (64 QAM is optional at the UE side) [3]. We notice also that the MIMO (Multiple Input Multiple Output) technique is also supported in LTE with two transmit antennas being the basic configuration.

The physical layer of LTE is identified as the design principle of radio resources usage based on the fact that these resources are shared among all the UEs and also dynamically allocated. Indeed as shown in Figure 2.3, the radio resources in LTE networks are assigned in the time and the frequency domain. In the time domain, the

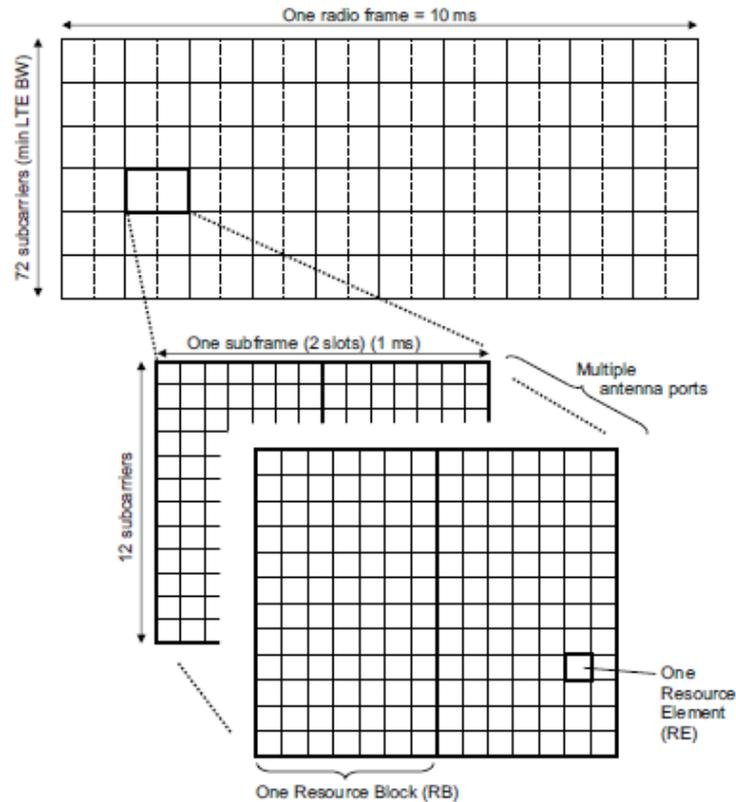


FIGURE 2.3: LTE resource structure [10]

radio resources are distributed every Transmission Time Interval (TTI) which lasts 1 ms [19]. We point out that 10 consecutive TTIs form a LTE frame. Besides, each TTI (subframe) is composed of two time slots with length of 0.5 ms and corresponding to 7 OFDM symbols (short cyclic prefix configuration). In the frequency domain, instead, the total bandwidth is divided into subchannels of 180 kHz. Each subchannel comprises 12 consecutive subcarriers which have a constant spacing of 15 kHz.

A time/frequency radio resource covering one slot in the time domain and one subchannel (180 kHz) in the frequency domain forms a Resource Block (RB) and corresponds to the smallest radio resource unit which can be assigned to a UE during the scheduling process. Radio resources can be also expressed in terms of Physical Resource Block (PRB) which cover instead one TTI (2 slots) in the time domain and one subchannel in the frequency domain. According to the system bandwidth configuration, the number of RBs can vary in LTE. Table 2.3 lists the number of RBs for different types of LTE bandwidth.

As seen above, each radio interface layer processes data and control messages according to its functions and convey the processed informations flows to the next layer. They communicate via different types of *channels* that are differentiated by the kind of

TABLE 2.3: LTE system bandwidth configuration

Channel bandwidth [Mhz]	1.4	3	5	10	15	20
Number of resources blocks	6	15	25	50	75	100

information they carry and by the way in which the information flow are processed. In the part below, we give more informations about these communications channels.

2.4.2 Communication channels

As communication channels, LTE defines the logical channels, the transport channels and the physical channels. We detail each of them in the parts below.

2.4.2.1 Logical channels

The logical channels define *what type* of information is transferred. They are used to carry data and control messages exchanged between the RLC layer and the MAC layer. So, they are classified in two ways. We have the logical traffic channels used to carry data in the user plane and the logical control channels used to carry the control messages in the control plane. Furthermore the logical control channels can be either dedicated control channels or common control channels. Table 2.4 presents the logical channels used in LTE.

TABLE 2.4: Logical channels used in LTE

Channel name	Acronym	Type
Broadcast Control Channel	BCCH	Control channel
Paging Control Channel	PCCH	Control channel
Common Control Channel	CCCH	Control channel
Dedicated Control Channel	DCCH	Control channel
Multicast Control Channel	MCCH	Control channel
Dedicated Traffic Channel	DTCH	Traffic channel
Multicast Traffic Channel	MTCH	Traffic channel

2.4.2.2 Transport channels

The transport channels define *how* and *with what type of characteristics* the information flow is transferred over the air interface. They are used to carry data and control messages exchanged between the MAC layer and the Physical layer. The transport channels are differentiated by the ways in which they are used. Table 2.5 lists the transport channels used in LTE.

TABLE 2.5: Transport channels used in LTE

Channel name	Acronym	Direction
Broadcast Channel	BCH	Downlink
Downlink Shared Channel	DL-SCH	Downlink
Paging Channel	PCH	Downlink
Multicast Channel	MCH	Downlink
Uplink Shared Channel	UL-SCH	Uplink
Random Access Channel	RACH	Uplink

Besides the transport channels, different types of control informations are used to support several physical layer operations. Here are the most important:

Downlink Control Information (DCI) : The DCI conveys all information related to the scheduling decisions (for both uplink and downlink), Modulation and Coding Schemes (MCS) and the transmitter power control (TPC).

Control Format Indicator (CFI) : CFI reveals the number of symbols used for the DCI in the subframe.

HARQ Indicator (HI) : It is used to carry the HARQ ACK/NACK in response to an uplink transmission.

Uplink Control Information (UCI) : The UCI is used to convey the Scheduling Requests (SR) of the UE and the HARQ ACK/NACK in response to a downlink transmission.

2.4.2.3 Physical channels

The physical channels define *where* the information flow is transmitted over the air interface. They are used by the physical layer to transmit data and control messages on the air interface. The physical channels are also differentiated by the ways in which they are used and are generally grouped in two parts, namely physical data channels and physical control channels. Table 2.6 lists the physical channels used in LTE [19].

TABLE 2.6: Physical channels used in LTE

Channel name	Acronym	Type	Direction
Physical Downlink Shared Channel	PDSCH	Data channel	Downlink
Physical Broadcast Channel	PBCH	Data channel	Downlink
Physical Multicast Channel	PMCH	Data channel	Downlink
Physical Uplink Shared Channel	PUSCH	Data channel	Uplink
Physical Random Access Channel	PRACH	Data channel	Uplink
Physical Control Format Indicator Channel	PCFICH	Control channel	Downlink
Physical HARQ Indicator Channel	PHICH	Control channel	Downlink
Physical Downlink Control Channel	PDCCH	Control channel	Downlink
Physical Uplink Control Channel	PUCCH	Control channel	Uplink

It results from the different channels description that there is a link between these channels, based on the purpose and the content. It necessitates mappings between the different channels, namely appropriate mappings between the logical channels and transport channels and suitable mappings between transport channels and physical channels. Figure 2.4 and Figure 2.5 present the standardized mappings between the different channels and between control information and physical channels, respectively.

As seen throughout this section, the LTE radio protocols provide the means for setting up, reconfiguring and releasing the radio bearer which are used to transfer applications data or control informations between the UE and the network. In order to help applications data and control informations to be transported across the LTE radio interface, different types of channels are used. These channels are mapped in an appropriate way in order to separate the different types of information flow and allow them to be transported across the radio access network in a suitable fashion. This is in favour of the aim of LTE which is to efficiently support various QoS classes of services.

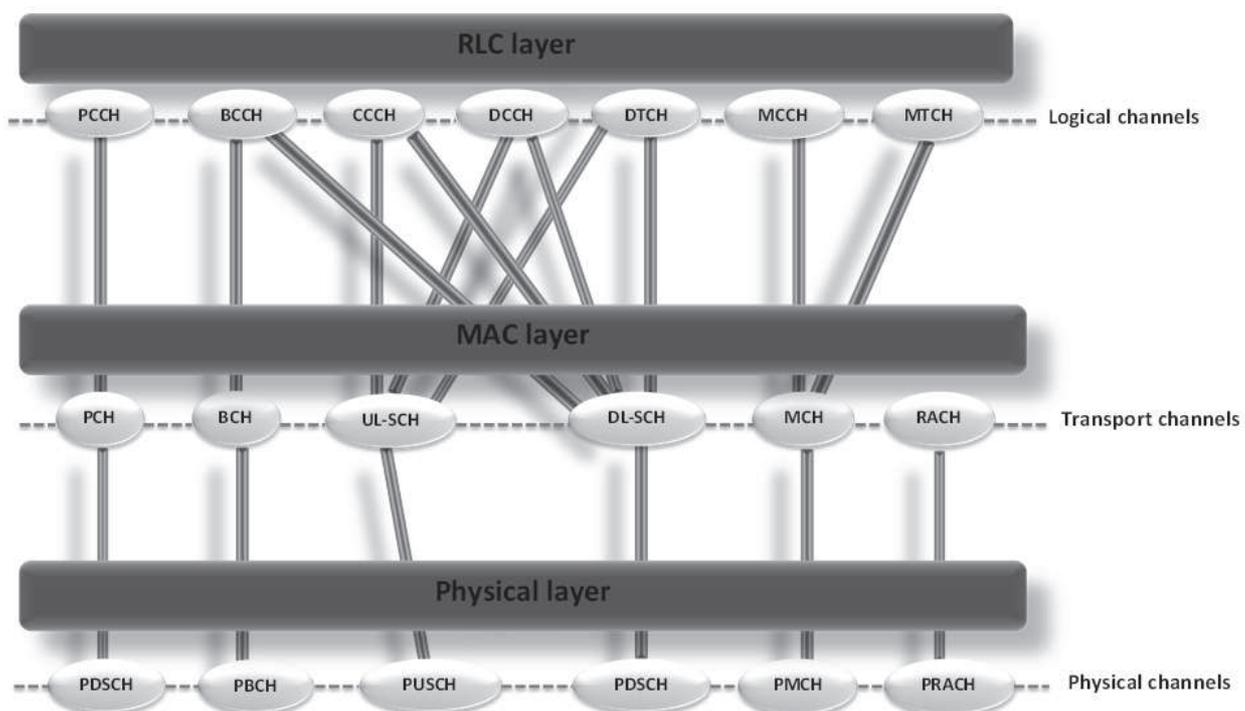


FIGURE 2.4: Mapping between the different types of channels [12]

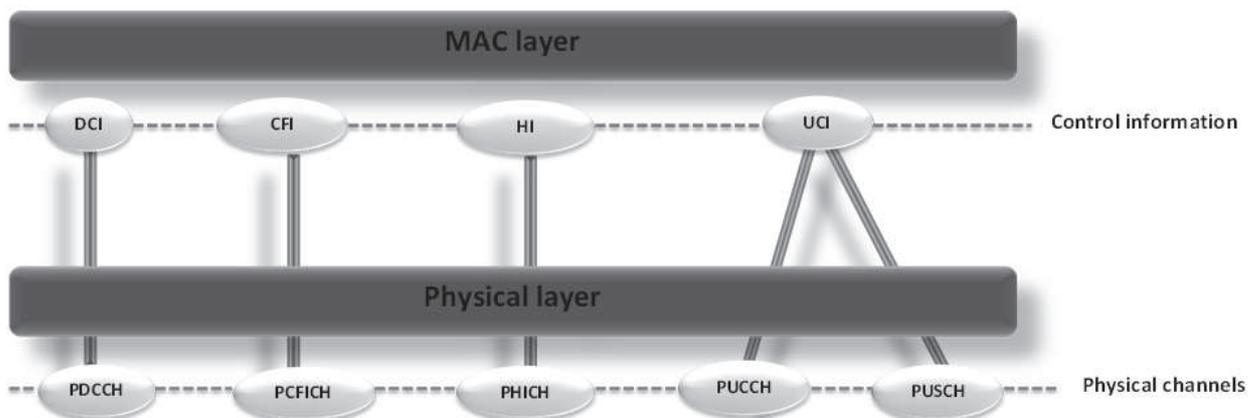


FIGURE 2.5: Mapping of control information to physical channels [12]

Nethertheless, LTE networks use several other techniques in addition to the aforementioned to support various class of services with strong QoS support. Among all these innovative concepts, the most important is the Radio Resource Management (RRM) techniques. Indeed, the RRM function aims at using the radio resources in an efficient way and serving the users according to their configured QoS parameters. The next

section will give more details about these RRM techniques.

2.5 Radio resources management procedures

As mentioned at the end of the previous section, RRM ensures that the radio resources are used in an efficient fashion and also serves the UEs according to their configured QoS parameters. To reach its goals, it includes a wide range of techniques and procedures such as strategies and algorithms which aims at utilizing the limited resources and radio network infrastructure as efficiently as possible. However in this section, we are interested in reporting mechanisms that support radio resources scheduling in the LTE network, namely the Channel Quality Indicator (CQI), the radio link adaptation, the HARQ operation and the TPC. The parts below will give more details.

2.5.1 CQI

The role of the CQI manager is to process the received CQI reports and the Sounding Reference Signals (SRSs) from the UEs attached to the eNB. The CQI is used on one hand for scheduling decisions and on the other hand, it provides the eNB information about the link adaptation parameters that the UEs can support.

Indeed, in the downlink the eNB uses the CQI feedback received from the UEs as an indication of the data rate which can be supported by the downlink channel. In the uplink, the eNB makes directly its own estimate of the data rate which can be supported using the SRS sent by the UE. As a consequence, by exploiting the CQI feedbacks and the SRS, the eNB can, not only select the best Modulation and Coding Schemes (MCS) for an efficient downlink or uplink transmission, but also have a view on the UEs experiencing the best channel conditions and give them priority when processing the scheduling.

2.5.2 HARQ

The HARQ procedure is very useful for errors-sensitive applications since it is used for retransmissions. It is based on the use of the Stop and Wait (SAW) algorithm and performed by eNB and UE through the exchange of ACK/NACK messages [20].

Indeed, when a transmitted packet is unsuccessfully decoded by the receiver, it sends a NACK message that will indicate to the sender to perform a retransmission by sending the same copy of the erroneous packet. Once the packet is retransmitted, the receiver

will try to decode it combining the retransmitted version with the original version, and send an ACK message in case of a successfully decoding.

However, the HARQ process also permits to the receiver to correct minor errors without requesting a new message from the sender.

2.5.3 AMC and TPC

As mentioned above, the CQI are used for link adaptation and scheduling decisions. Indeed, based on CQI feedbacks and the SRS, the AMC module chooses the appropriate MCS trying to maximize the supported throughput with a given target Block Error Rate (BLER) [12]. As consequences, UEs with best channel conditions will achieve higher data rates, whereas those experiencing bad channel conditions, will still have lower throughput.

Due to the limited number of the admitted MCS, the system throughput is upper-bounded. Consequently, over a certain threshold, there will not have any throughput gain if the Signal to Interference plus Noise Ratio (SINR) increases. In this way, the AMC module works in coordination with the TPC module. The TPC module consists of various mechanisms which aim at adjusting dynamically the UE transmission power on the radio link to face the changing nature of the instantaneous channel conditions. Thus, it allows to save the UE battery energy while allowing various level of bitrate to be achieved [11].

As seen throughout this section, RRM ensures that the radio resources are used in an efficient fashion and also serves the UEs according to their configured QoS parameters. To achieve its goals, RRM encompasses a wide range of techniques and procedures but among all of these mechanisms, the fundamental is accomplished by the eNB through the resources *scheduling procedure*. In the following section, we will go to the heart of our topic and present in details this vital function which have been addressed throughout this thesis.

2.6 Scheduling

Scheduling is part of the crucial features of LTE networks and plays a vital role since it is responsible of efficiently allocating the radio resources. Indeed, the effective use of radio resources will lead to reach the systems targets and satisfy the QoS requirements of the UEs. The scheduling function is handled by the eNB that decides every TTI which resources will be allocated to the UEs. The resources to allocate are expressed in terms

of RBs or PRBs. The RBs span over one slot in the time domain and one subchannel in the frequency domain, whereas the PRBs cover one TTI in the time domain and one subchannel in the frequency domain. They correspond to the smallest radio resource unit which can be assigned [21].

The scheduling process is usually based on the comparison of per-RB metrics. It means that the eNB takes the scheduling decisions after comparing the RBs metrics of all the UEs. More precisely, the n -th RB of the system can be allocated to the user z , if and only if its metric $m_{z,n}$ is the highest one among the other metrics. Mathematically it can be expressed as:

$$m_{z,n} = \max_i \{m_{i,n}\} \quad (2.1)$$

We notice that the expression of the scheduling metric can vary depending on the desired performance to reach. We list below the main parameters used by the eNB to compute the scheduling metric:

- **Channel quality:** As explained in the RRM section (part 2.5.1), the CQI is used by the eNB to have an indication of the UEs channel conditions. By exploiting the CQI feedbacks and the SRS, the eNB can have a view on the UEs experiencing the best channel conditions and give them priority when processing the scheduling;
- **QoS parameters:** As explained in the QoS management section (part 2.3), whenever a UE runs an application, a bearer is set up in the system. Each bearer has a QoS profile, namely the QCI, which includes a set of QoS parameters depending on the application data. The eNB, in order to satisfy the QoS requirements of the UEs, can use the QCI values associated to their flows to give them priority at the scheduling process;
- **Resources allocation history:** Fairness is an important aspect in LTE networks. Indeed, all the UEs including the cell-edge users must be served. For this purpose, when processing the scheduling the eNB can use information about the past achieved performance of the UEs in order to improve fairness;

In this section, as said before we go to the heart of our topic. At first, we describe the different steps of the scheduling process in downlink as well as in the uplink. Then, the key issues regarding the design of a LTE scheduler are addressed. Finally, a detailed classification of the most known and used strategies and algorithms is provided for both downlink and uplink.

2.6.1 Scheduling process

The scheduling process is slightly different according to whether we are in uplink or downlink.

2.6.1.1 Scheduling process in the downlink

In the downlink direction, the eNB allocates the RBs directly to the flows, instead of UEs. For instance, a UE running different applications will receive scheduling decisions for each of its flows instead of one scheduling decision for the whole flows. As said before, the scheduler works in general with the granularity of one TTI. This means that during the scheduling process, the eNB executes a list of operations which is repeated and renewed in general every TTI [22].

At first the eNB, which is obviously aware of the amount of data awaiting in its buffer, prepares the list of the flows which could be scheduled and allocated RBs in the current TTI. Then, the different UEs report their CQI. As explained previously, the CQI feedbacks will be very useful to the eNB. Firstly they will allow the eNB to estimate the UEs downlink channel quality and to have an indication of the data rate which will be supported by the downlink channel. Secondly the CQI feedbacks will be used as input parameters in the scheduling metric.

At the second step, after processing the CQI feedbacks, the eNB computes for each RB belonging to the TTI, the scheduling metrics of each candidate flow according to the defined scheduling policy. Then, each RB is allocated to the flow presenting the highest metric on the concerned RB, in accordance with equation 2.1. Afterwards, the eNB calculates for each scheduled flow the amount of data which will be transmitted. For this purpose, it uses the AMC module which chooses the appropriate MCS trying to maximize the supported throughput with a given target BLER.

At the third step, the eNB conveys the scheduling decisions to the UEs on the PDCCH. The PDCCH carries the DCIs which contain all information which is necessary for the UEs to be able to identify the resources on the PDSCH and decode them.

Finally, Each UE reads the PDCCH payload and if resources have been allocated, it accesses the part of shared channel containing its data.

A general view of the process is illustrated in Figure 2.6

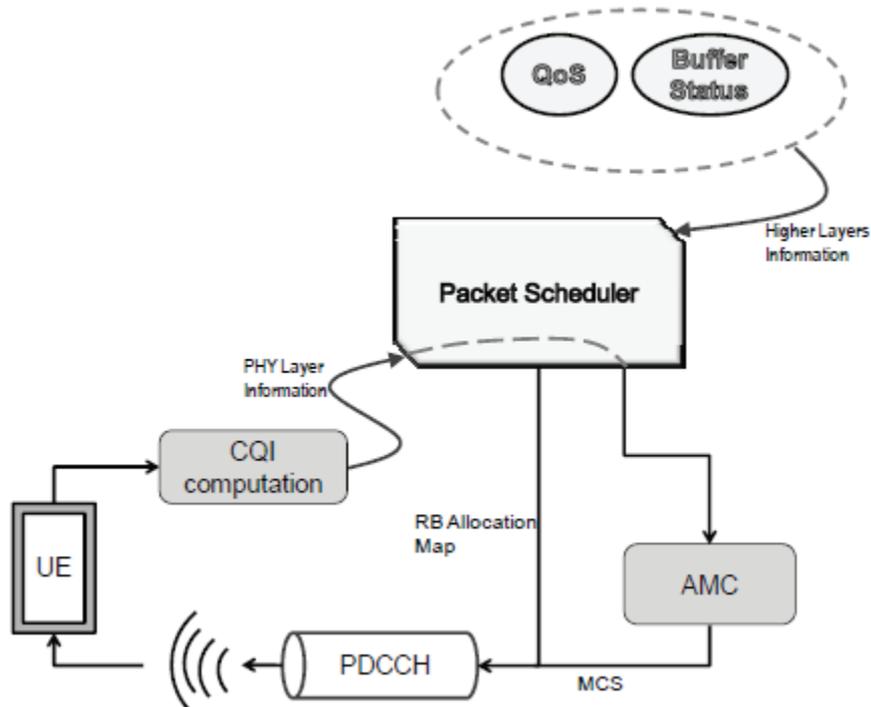


FIGURE 2.6: General model of a downlink packet scheduler [11]

2.6.1.2 Scheduling process in the uplink

As explained in Chapter 1, the use of SC-FDMA in the uplink imposes several constraints to be taken into account. The scheduling process is also concerned. Indeed, contrary to downlink, the resources allocation is done per UEs in the uplink, even though each UE may have several flows [23]. In other words, a UE running different applications will receive one scheduling decision for the whole flows in the considered TTI. In addition, RBs to allocate to a scheduled UE must be contiguous. Besides, the scheduling procedure in the uplink behaves differently since the eNB at the beginning of the process has no information about the amount of the UEs data. Indeed, in the downlink the eNB is obviously aware of the amount of UEs data awaiting in its buffer but in the uplink it is different. It is incumbent upon the UEs to inform the eNB about the amount of buffered data to be transmitted and their priority.

At the beginning of the process, since the eNB has no information about the amount of the UEs data, it needs to be informed by them. For this purpose, every UE having data to send in their buffer transmits a Scheduling Request (SR) over the PUCCH. This lets the eNB know that data is awaiting in the UE buffer. However the SR is just a signal sent to warn the eNB and does not contain any information about the amount of data to transmit. Thus, the eNB in order to be aware of this amount and allow transmission, sends a minimal grant back to the UE over the PDCCH. Like in the downlink, the

PDCCH carries the UCIs which contain all information which is necessary for the UEs to be able to identify the resources on the PUSCH and use them for transmissions. After receiving the minimal grant, the UE carries out its first transmission over the PUSCH. In this first data transmission, the UE includes the Buffer Status Report (BSR) which gives an indication on the amount of data awaiting in its buffer. The scheduling process is then performed based on the BSR and this BSR will help the eNB to make more accurate decisions about the amount of radio resources to grant in future TTIs [24]. A general view of the interaction between the UE and the eNB in the scheduling process is depicted in Figure 2.7

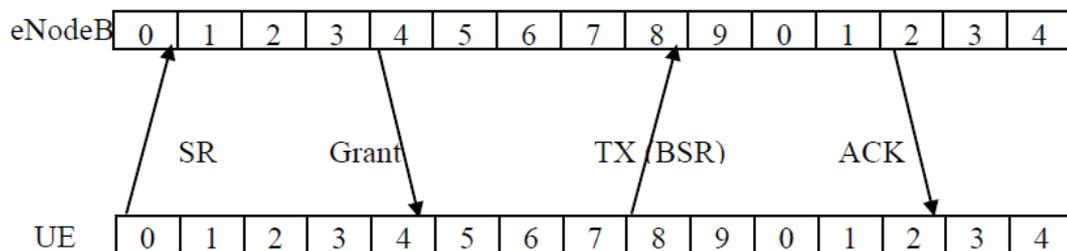


FIGURE 2.7: Interactions between eNB and UE in uplink scheduling [24]

After describing the different steps of the scheduling process, we will present the different types of allocation strategies that have been designed for LTE systems. Nevertheless, it is important to begin with the factors which influence the definition of such resources allocation strategies.

2.6.2 Key design aspects

Designing effective scheduling strategies supposes the conception of methods and schemes which meet the required expectations and deal with several other constraints. In other words, an efficient scheduling strategy should do a trade-off between all these specificities. We present below a list of the main design elements which have an undeniable effect on the conception of an allocation strategy for LTE.

1. **Spectral efficiency:** Better spectral efficiency is one of the main goals to be realized within LTE. Indeed as mentioned in the introductory part of this Chapter, LTE targets a spectral efficiency 2 - 4 times better than HSPA systems. In other words resources must not be wasted. It supposes to maximize the number of UEs served in a given TTI or to always serve the UEs experiencing the best channel conditions.

2. **Fairness:** The fact of always serving the UEs experiencing the best channel conditions will certainly enable effective channel utilization in terms of spectral efficiency, but at the cost of starving those experiencing the worst channel conditions. Clearly, it will lead to very unfair resources sharing among the UEs. However, fairness is an important aspect and a major requirement in LTE networks. As consequence, all the UEs including the cell-edge users must be served.
3. **QoS Provisioning:** QoS provisioning is also a very important aspect in next generations of mobile networks and especially for LTE networks which the main objective is to provide the best QoS supports to strict QoS constrained applications. As previously mentioned in sections 2.3 and 2.5, whenever a UE runs a QoS constrained application, a dedicated bearer is set up in the system. These dedicated bearers have a QoS profile which enables special RRM procedures. Thus, the scheduler, in order to satisfy the QoS requirements of the UEs, has to take into account the QCI values associated to their flows.
4. **Uplink constraints:** Contrary to the downlink, the uplink radio access scheme, namely SC-FDMA allows the UEs to only transmit their data in a single carrier mode. Consequently, the scheduler has the constraint to allocate only contiguous RBs to the scheduled UEs.

2.6.3 Scheduling strategies classification

In this part, we classify and detail different resources allocation algorithms introduced for LTE systems. They are well-known schemes and widely used in the literature.

There are two major categories of resources allocation strategies in LTE. They are known in the literature as *channel-unaware scheduling* and *channel-aware scheduling* [25]. Furthermore, the channel-aware schedulers can be separated on the basis of QoS support as *QoS-unaware* and *QoS-aware* schedulers. In addition to these schemes, semi-persistent and persistent solutions are also used. For an ease of presentation we divide them into four groups, namely channel-unaware strategies, channel-aware/QoS-unaware strategies, channel-aware/QoS-aware strategies and semi-persistent/persistent strategies.

2.6.3.1 Channel-unaware strategies

These types of strategy have been proposed to improve fairness among UEs and also to prevent the delivery of packets after their delay deadline [26]. In LTE networks they are used for the most part in conjunction with the channel-aware schemes. We detail below the well-known strategies.

- 1) **First in First out (FIFO)** : As indicated by its acronym meaning, the FIFO scheme allocates RBs to the UEs depending on the time instant when their requests were issued. The scheduling metric of the z -th user on the n -th RB can be expressed as:

$$m_{z,n}^{FIFO} = t - T_z \quad (2.2)$$

where t is the current time and T_z is **the time instant** when the request was issued by the z -th user.

- 2) **Round Robin (RR)** : RR algorithm is a starvation-free RBs method. With this algorithm, all the candidate users are firstly ordered randomly in a queue. After that, the RBs are allocated to the UE which is at the front of the queue. After receiving RBs, this UE is put at the rear of the queue and the rest of steps follow the same way, until no UE asks for RBs. The scheduling metric is expressed as :

$$m_{z,n}^{RR} = t - T_z \quad (2.3)$$

where T_z refers to the **last time** when the z -th user was served.

- 3) **Blind Equal Throughput (BET)** : This strategy is used to provide fairness among the UEs in terms of throughput. For this purpose, the BET algorithm considers the past average throughput of the candidate UEs [27]. The scheduling metric is expressed as:

$$m_{z,n}^{BET} = 1/\overline{R^z}(t-1) \quad (2.4)$$

with

$$\overline{R^z}(t) = \beta \overline{R^z}(t-1) + (1-\beta)r^z(t) \quad (2.5)$$

where $0 \leq \beta \leq 1$, $\overline{R^z}(t)$ is the past average throughput experienced by the z -th user at time t and $r^z(t)$ the data rate achieved by the z -th user at time t .

From equation 2.4, it can be seen that the BET algorithm gives priority to the UEs having lower average throughput in the past. Consequently, the UEs with bad channel conditions are ensured to have more resources.

- 4) **Largest Weighted Delay First (LWDF)** : LWDF algorithm is used to avoid deadline expiration of the Real Time (RT) applications packets [28]. Indeed, RT applications require that each packet must be delivered within a certain deadline to avoid packet drops. The scheduling metric is expressed as:

$$m_{z,n}^{LWDF} = \alpha_z \cdot D_{HOL,z} \quad (2.6)$$

where α_z is given by:

$$\alpha_z = -\frac{\log \delta_z}{\tau_z} \quad (2.7)$$

$D_{HOL,z}$ (Head Of Line Delay) is the delay of the first packet to be transmitted by the z -th user.

δ_z represents the acceptable probability for the z -th user that a packet is dropped due to deadline expiration. As for τ_z , it is the delay threshold for the z -th user.

As seen, these types of scheme improve fairness among UEs and also avoid deadline expiration of RT applications packets. Nevertheless they are not suitable for LTE networks since their allocation policies do not include a parameter that takes into account the variant nature of the radio channel conditions.

2.6.3.2 Channel-aware/QoS-unaware strategies

Channel-aware scheduling strategies allocate RBs by taking into account the channel conditions of the candidate UEs. To have an indication about these channel conditions, the eNB exploits the SRS and the CQI feedbacks received from the UEs. They will help the eNB to have an estimate of the data rate which could be supported by the downlink and the uplink channel. We present below the well-known schemes.

1) Maximum Throughput (MT) : MT algorithm aims at increasing the spectral efficiency of the network. For this purpose, only UEs experiencing the best channel conditions are given priority. The scheduler first exploits the SRS and the CQI feedbacks received from the UEs to have an estimate on the data rate which can be achieved by the UEs on the different subchannels. Then it allocates the RBs to the UEs which can achieve the highest throughput on the considered subchannel [29]. The scheduling metric is computed as:

$$m_{z,n}^{MT} = d_n^z(t) \quad (2.8)$$

where $d_n^z(t)$ is the expected data rate for the z -th user at time t on the n -th RB. It can be calculated as:

$$d_n^z(t) = \log_2[1 + SINR_n^z(t)] \quad (2.9)$$

MT permits to maximize the cell throughput, but at the cost of starving the UEs with the worst channel conditions.

2) Proportional Fair (PF) : The PF scheme has been designed specifically for the downlink direction and for the Non-Real Time (NRT) applications. PF algorithm performs a trade-off between fairness and spectral efficiency. Indeed, it improves fairness in terms of throughput among the UEs without losing the spectral efficiency [30]. The scheduling metric is a combination of the BET metric and MT metric. It is expressed as:

$$m_{z,n}^{PF} = m_{z,n}^{MT} \cdot m_{z,n}^{BET} = d_n^z(t) / \overline{R^z}(t-1) \quad (2.10)$$

From the scheduling metric expression, it can be seen that the past average throughput is used as a weighting factor of the expected data rate. In this condition, the UEs having the worst channel conditions are ensured to be served within a certain amount of time. The parameter β in equation 2.5 is an important parameter since it allows to define to the time window T_f , over which fairness will be enabled. Indeed the relation between β and T_f is expressed as:

$$T_f = 1/(1 - \beta) \quad (2.11)$$

PF scheme has been modified to be adapted for uplink system due to the fact related to the contiguous RB allocation required by SC-FDMA. Two promising strategies, namely First Maximum Expansion (FME) and Recursive Maximum Expansion (RME) have been proposed in [31] and [32] and are based on the PF utility function. FME algorithm allocates the RBs to a UE beginning from the RB where the UE has the highest metric value in the resources allocation matrix and expand the assignment on the right and the left sides of this RB. The resources allocation of the UE is stopped whenever the algorithm finds another user having a higher metric on the considered RB. RME acts like FME scheme but with the particularity of performing a recursive search of the maximum metric.

Taking into account the radio channel quality and the fairness aspect are very important in mobile networks since it permits not only to achieve high performance such as higher throughput but also to guarantee services to the cell edge users. However as explained in section 2.6.2, spectral efficiency and fairness are not the unique targets of

LTE. Indeed, LTE networks has to provide the best QoS support to the QoS constrained applications. The strategies detailed below, take into account this aspect.

2.6.3.3 Channel-aware/QoS-aware strategies

QoS support to QoS constrained flows is one the main objectives of LTE networks. QoS support is managed in LTE by associating a set of QoS parameters, namely the QCI and the ARP, to each flow. Knowing these QoS parameters and their required values, the scheduler can take them into account in the scheduling metric and thus fulfill the QoS requirement of the different UEs. This part gives details on schedulers which aim at guaranteeing bounded delay, while at the same trying to increase the spectral efficiency and improve fairness.

1) Modified Largest Weighted Delay First (M-LWDF) : This algorithm was presented in [29] and aims at increasing the QoS support to RT flows by delivering the packets before the delay deadline. For this purpose, the radio channel conditions and the state of the queue with respect to delay constraint are taken into account in the scheduling policy. Furthermore, due to the fact that there can be NRT flows in the cell, they are handled differently. PF algorithm is used for the NRT flows and the RT flows are treated with the following metric:

$$m_{z,n}^{MLWDF} = \alpha_z \cdot D_{HOL,z} \cdot m_{z,n}^{PF} = \alpha_z \cdot D_{HOL,z} \cdot \frac{d_n^z(t)}{R^z(t-1)} \quad (2.12)$$

where $D_{HOL,z}$ is the delay of the head of line packet and α_z is calculated as in equation 2.7.

2) EXPonential/Proportional Fair (EXP/PF) : EXP/PF algorithm is the QoS-aware extension of the PF scheme. It is used to give priority to the RT flows which have their HOL packet delays approaching the delay deadline. Like the M-LWDF scheme, EXP/PF strategy can handle RT and NRT flows. The NRT flows are treated using the PF metric and the RT flows using the following metric [33]:

$$m_{z,n}^{EXP/PF} = \exp\left(\frac{\alpha_z \cdot D_{HOL,z} - \chi}{1 + \sqrt{\chi}}\right) \cdot \frac{d_n^z(t)}{R^z(t-1)} \quad (2.13)$$

where

$$\chi = \frac{1}{N_{rt}} \cdot \sum_{z=1}^{N_{rt}} \alpha_z \cdot D_{HOL,z} \quad (2.14)$$

and N_{rt} is the number of active RT flows.

2) LOG and EXP rules : These strategies have been proposed in [34] and are also sensitive to radio channel conditions and the head of line packet delay of RT flows. For the LOG rule scheme, the scheduling metric is computed as:

$$m_{z,n}^{LOGrule} = b_z \cdot \log(c + a_z \cdot D_{HOL,z}) \cdot \Gamma_n^z \quad (2.15)$$

where b_z , c and a_z are tunable parameters and Γ_n^z represents the spectral efficiency for the z -th user on the n -th user subchannel. Moreover the optimal performance is achieved using $a_z = 5/(0.99 \cdot \tau_z)$, $b_z = 1/E[\Gamma^z]$ and $c = 1.1$

As for the EXP rule, it is more robust and is an improvement of EXP/PF strategy. Its scheduling policy is expressed as:

$$m_{z,n}^{EXPrule} = b_z \cdot \exp\left(\frac{a_z \cdot D_{HOL,z}}{c + \sqrt{(1/N_{rt}) \cdot \sum_t D_{HOL,t}}}\right) \cdot \Gamma_n^z \quad (2.16)$$

where the optimal parameters are :

$$\begin{cases} a_z \in [5/(0.99 \cdot \tau_z), 10/(0.99 \cdot \tau_z)] \\ b_z = 1/E[\Gamma^z] \\ c = 1 \end{cases}$$

It can be seen from the scheduling metric expressions that EXP rule algorithm is more efficient than LOG rule scheme since the EXP rule scheduling policy takes into account the head of line packet delays of all the RT flows. Besides, the exponential function has the characteristic to grow faster than the logarithm function.

2.6.3.4 Persistent and semi-persistent strategies

As mentioned in section 2.6.1.1, the eNB executes during the scheduling process a list of operations which is repeated and renewed in general every TTI. This dynamic aspect of the scheduling process has the particularity of taking advantage of the multi-user diversity gain, but at the cost of increasing the control overheads generated during the process. Indeed, during the scheduling process, the eNB has to inform the UEs about its scheduling decisions. For this purpose it sends DCIs and UCIs messages to notify them. In case of high traffic load, the radio resources reserved for control informations can become insufficient and lead to the degradation of the QoS provisioning capabilities [35].

To avoid the degradation of the QoS provisioning capabilities, persistent strategies have been proposed. With this algorithm, the eNB does not have to inform the scheduled

UEs every TTI [36]. Indeed, with this strategy the eNB allocates the same RBs to the UEs in a permanent mode [37]. Once the UEs have been notified, they will know in advance the part of the shared channel where they should decode their data (downlink) or perform their data transmissions (uplink) with no necessity of receiving any additional DCI or UCI.

In addition to the fact that persistent strategies do not profit from the radio channel variations, the HARQ operations are not supported. However retransmissions are important in LTE since a unique transmission of packets is insufficient to guarantee their decoding at the receiver node side. For this purpose, the semi-persistent schemes have been introduced. They perform a trade-off between dynamic scheduling scheme and persistent strategy and are most of the time used for VoIP applications. For these types of flows, the semi-persistent strategy increases their capacity by maximising the number of supported VoIP calls [38].

As seen throughout this section, the scheduling process plays a major role in LTE networks since it is responsible of efficiently allocating the radio resources. Indeed, by performing a correct resources allocation, the system targets will be met and the QoS requirements of the UEs will be satisfied. We have also seen that the scheduling process consists in designing cutting-edge algorithms that meet the required expectations. In a practical way, from our presentation of the existing LTE scheduling strategies, it results that an efficient scheduling algorithm must do in general a trade-off between fairness, spectral efficiency and QoS requirements.

2.7 Conclusion

The emergence of new applications has led to a growing demand for networks services with strict constraints on delays and bandwidth requirements. To face this growing demand for packet-based services, the 3GPP has introduced the LTE specifications.

As seen in this chapter, LTE is a very promising technology with respect to previous generation of mobile networks, as it is based on an all-IP architecture that aims at supporting several high quality services such as video streaming, VoIP, videotelephony and online gaming. In order to fulfill these ambitious goals, LTE introduced new technical principles and concepts. They encompass in addition to the new network architecture, the use of a new QoS management policy through the bearer management, the use of new multiple access schemes such as OFDMA in the downlink and SC-FDMA in the uplink, and a series of mechanisms at different layers of the protocol stack, able to efficiently exploit the wireless link capacity up to the Shannon limit.

The most important RRM mechanism is performed through the scheduling process which consists in designing efficient algorithms which take into account both the QoS requirements and the physical constraints for distributing radio resources among users or flows. This process represents a great challenge since the proposed strategy must be, not only a simple and reliable algorithm which achieves a trade-off between various aspects such as fairness criteria, overall cell throughput maximization and QoS requirements, but also a scheme which deals with the fact related to the contiguous RBs allocation required by SC-FDMA. Several algorithms dealing with these issues have been designed and presented within this chapter. However, to the best of our knowledge, the issue related to the scheduling of radio resources in a high mobility context in LTE remains to be addressed.

For those reasons, we have conducted this thesis with the intention of proposing effective strategies and methods aiming at improving resources allocation in such scenarios. Nevertheless, as pointed in Chapter 1, we will focus the study initially on a context where the high mobility of the user is not taken into account. This will help not only to reduce the complexity of the work but also to critically analyze the existing works and propose new solutions to improve QoS of real time applications. Once these tasks done, the high mobility parameter will be taken into account and also suitable methods dealing with this context will be proposed.

In the two following chapters, we will present our innovative strategies designed for improving resources allocation in on a context where the high mobility of the user is not taken into account. Issues related to the downlink direction will be first addressed.

Chapter 3

Resources allocation in the Downlink direction

This chapter presents our two major contributions designed to improve downlink resources allocation of LTE networks in a context where the high mobility of the users is not taken into account. The first one is an innovative scheme which ameliorates resources assignment, particularly in overbooking scenarios by doing a trade-off between fairness, overall system throughput and QoS requirements. The second one is an enhanced scheduling scheme which provides strict delay bounds and guarantees very low packet loss rate to multimedia services. The performance of the proposed schemes have been evaluated by simulations and compared to other schemes of the literature. The analyses demonstrated the effectiveness of our strategies and showed that they outperformed the other methods.

3.1 Introduction

As stated in the previous chapters, LTE represents an emerging and promising technology which has been designed to face the growing demand for network services by providing higher data rates, low latency and improved spectral efficiency. Scheduling plays a major role to achieve these goals because it is responsible for resources allocation. Indeed, an effective allocation will help to meet the system performance targets and satisfy the QoS requirements of the services. In this way, we aim in this thesis at designing efficient resources allocation strategies in order to improve QoS of real time applications in a context where the users have very high velocities. We explained in Chapter 1 that the study will be carried out in two steps in order to reach the required goal. More precisely in order to have an expert knowledge of the key facets of LTE scheduling, the

study will be conducted at first in a context where the high mobility aspect of the node is not taken into account. This will help not only to reduce the complexity of the work but also to critically analyze the existing works and propose new solutions to improve QoS of real time applications. Once this task done, the high mobility parameter will be added and also appropriate methods dealing with this context will be developed for QoS enhancement of real time applications. Nevertheless due to the existing differences between the uplink and the downlink direction, appropriate scheduling schemes will be also designed in each of the aforementioned directions.

Thus following this guideline, we address firstly in this chapter the problem of improving the scheduling of downlink communications in a context where the high mobility of the users is not taken into account. We begin with the downlink direction because the conception of efficient scheduling algorithms in this direction is less complicated compared to the uplink direction. Indeed in the uplink direction the proposed methods should not only deal with the constraints related to the scheduling process, but should also be compatible with the constraints set by this direction. In this way, by beginning with the downlink direction we will leverage the conception of effective scheduling algorithms in this direction to design appropriate ones for the uplink direction.

We have seen in the *state-of-the-art* part in Chapter 2 that the most important objective of the scheduling process in the downlink direction is to satisfy the QoS requirements of all users by trying to reach, at the same time, an optimal trade-off between spectral efficiency and fairness [4]. However, this goal becomes very challenging, especially in overbooking scenarios where we have very limited resources for a great number of users, and also in presence of RT multimedia applications which require strict constraints on packet delay and packet loss.

We present in this chapter our efficient downlink scheduling strategies which have been designed to deal with these two issues. The first one is a two-level scheduler scheme specifically adapted to the overbooking cases. The second one is an enhanced scheduling scheme used to improve the performance of multimedia services. Through the study of the existing techniques, we highlight their limitations and suggest our tailored solutions to meet the required needs.

3.2 A new two-level scheduling algorithm to face overbooking scenarios

3.2.1 Context

The requirement of new mobile networks with ubiquitous broadband access has led the 3GPP to define the LTE system that represents an important technology and brings new enhanced mechanisms with various integrated data services. Packets scheduling mechanism is the cornerstone of LTE systems because the best performance of the network can be reached by correctly allocating resources among users. This allocation must do a trade-off between various aspects such as spectral efficiency, QoS requirements and fairness. The problem is to find a simple and reliable algorithm which could achieve these objectives. However, the scheduling problem becomes more complex when resources are limited for the users, namely for overbooking scenarios. In fact as seen in Chapter 2, radio resources are not unlimited and their number in the system can be very restricted in reality. Therefore, these scenarios represent a great challenge for existing scheduling algorithms since it is to find a way to perform a good scheduling with a large number of users and very limited resources, while meeting fairness criteria, overall cell throughput maximization and QoS requirements.

For that reason, we propose a new scheduling algorithm for the downlink of LTE networks, which ensures a trade-off between all these specificities and improves resources assignment in overbooking scenarios, where the number of users is greater than the number of available resources [39].

3.2.2 Literature review

We have seen in the scheduling strategies classification in Chapter 2 (see 2.6.3) that several researches have been done to improve resources allocation in the downlink of LTE networks. However, only a few of these contributions addressed issues closely related to the overbooking cases. Here is a brief description of these techniques and their limitations.

In [11], a practical method based on PF scheme, is proposed to deal with fairness and spectral efficiency. It is a channel aware strategy in which the past average throughput of each user is used as a weighting element of the expected data rate. In that conditions, users with bad channel status are ensured to be served within a certain time. However, this strategy does not take into account the QoS requirements of the flows to schedule.

So it can become ineffective, especially in presence of multimedia flows which require strict constraints on packet delay and jitter.

This problem was solved in [33] where another efficient scheme has been proposed. The EXP/PF algorithm takes into account both the behaviour of a PF scheduler and of an exponential function of the end-to-end delay. It ensures a good balance between spectral efficiency, fairness and delay requirement. Although this method is very relevant, it does not focus precisely on overbooking scenarios (where the resources are limited for the users).

The scheduling strategy proposed in [39] focused on these scenarios and their contributions have been very relevant to our work. Indeed, their schemes focus on improving the resources allocation in that scenarios by providing fairness in the distribution of the RBs, while at the same keeping the system capacity utilization as high as possible. However, their methods do not take into account any QoS parameter to allocate resources. Thus, they cannot be efficient for real time flows in those scenarios. Our proposed scheme which is detailed below is more advanced and considers a QoS parameter for the resources allocation.

3.2.3 Key design aspects

Our method focuses on the overbooking scenarios and allows a good level of fairness between users in terms of RBs distribution, while keeping at the same time, as high as possible, the overall system capacity and performance of users having strict QoS requirements [40].

Our new scheduler *looks like* a two-level scheduler [4] which is in fact a joint Time and Frequency domain scheduler. At a first level, a Time Domain Packet Scheduler (TDPS) selects certain users among those connected to the base station and after, the RBs are assigned to these users by a Frequency Domain Packet Scheduler (FDPS) [41]. The final allocation decision results in the decision of the two schedulers that work in series. With this feature, our scheme helps to face the problem of overbooking scenarios where we have a great number of candidate users for few resources. Besides this aspect, different strategies are selected and applied to the schedulers in order to achieve a good trade-off between fairness, overall system capacity and QoS requirements. The parts below describe in details how our method has been designed.

3.2.3.1 Metric description

As mentioned previously, our strategy will be very helpful in overbooking scenarios. Let be u , r and c the number of users, the number of resources blocks and the allocation capacity of the system, respectively. The allocation capacity is given by:

$$c = \frac{r}{u} \quad (3.1)$$

The overbooking scenarios will occur if and only if:

$$c < 1 \quad (3.2)$$

The algorithm is designed to achieve a good trade-off between fairness, overall system capacity and QoS requirements. Let us explain in details how it performs.

- 1) Fairness :** Our method allows all users to be scheduled in a minimum of contiguous subframes. Let β be this number in terms of TTI. β represents the time window (the interval time) in which all the users are ensured to get a RB. β depends on u and r :

$$\beta \in \mathbb{N}, \beta \geq \frac{u}{r} \quad (3.3)$$

In case of $\lceil \frac{u}{r} \rceil$ is non-integer, the value of β is rounded to the first upper integer.

- 2) Throughput maximization based on QoS parameter :** Resources allocation for each terminal is usually based on the comparison of per-RB metrics. According to equation 2.1, a n -th RB can be allocated to a user z , if its metric $m_{z,n}$ is the highest one among the other metrics. In our case, the metric to use should allow to increase the overall system throughput and performance of users having strict QoS requirements. For that purpose, we define a metric which takes into account for each user its channel quality and the bearer priority of its flow [14]. This metric that we call the *decision index* will be used by the FDPS to allocate resources. It is given by:

$$m_{z,n}^{Proposed} = m_{z,n}^{MT} \cdot \frac{1}{BP_z} \quad (3.4)$$

where BP_z is the bearer priority of the flow of the user z . As explained in Chapter 2, bearers are created and set up whenever applications are running by the UE.

These bearers have a QoS profile which includes a set of QoS parameters depending on the application data for which the bearers have been set up. The bearer priority is part of these QoS parameters and is known by the scheduler. According to [14], the smaller is the value of the bearer priority, the higher is the priority level of the flow.

In this way, our method allows a number of users to be scheduled twice in β . Let γ this number.

$$\gamma \in \mathbb{N}, \gamma = (\beta \times r) - u \quad (3.5)$$

The set is composed of users having the γ best decision index.

3.2.3.2 Algorithm

Our method is implemented in these steps and the flowchart is depicted in Figure 3.1:

Step 1: Compute the time window β and perform all operations inside;

Step 2: Get the CQI feedbacks and the bearer priority of the flows;

Step 3: Compute the decision index of each flow;

Step 4: Look for flows which could be scheduled by ignoring those scheduled in the previous β ;

Step 5: Look for the highest value in the decision index set and the corresponding flow;

Step 6: Schedule this flow for the corresponding RB and ignore it for the next RBs;

while *All the flows have not been scheduled in β* **do**

Schedule the next flow with the highest decision index and ignore it for the next RBs;

end

if *Free RBs are left unused in β after scheduling all flows* **then**

Repeat from Step 5 by considering all flows;

else

Allocation is successfully completed in β ;

end

3.2.3.3 Model description

To model our scheduling strategy, we consider a LTE cell containing one eNB and many users generating different flows belonging to a particular traffic class. We assume m traffic classes which are used for QoS differentiation. $M = \{1, 2, 3, \dots, m\}$. In this

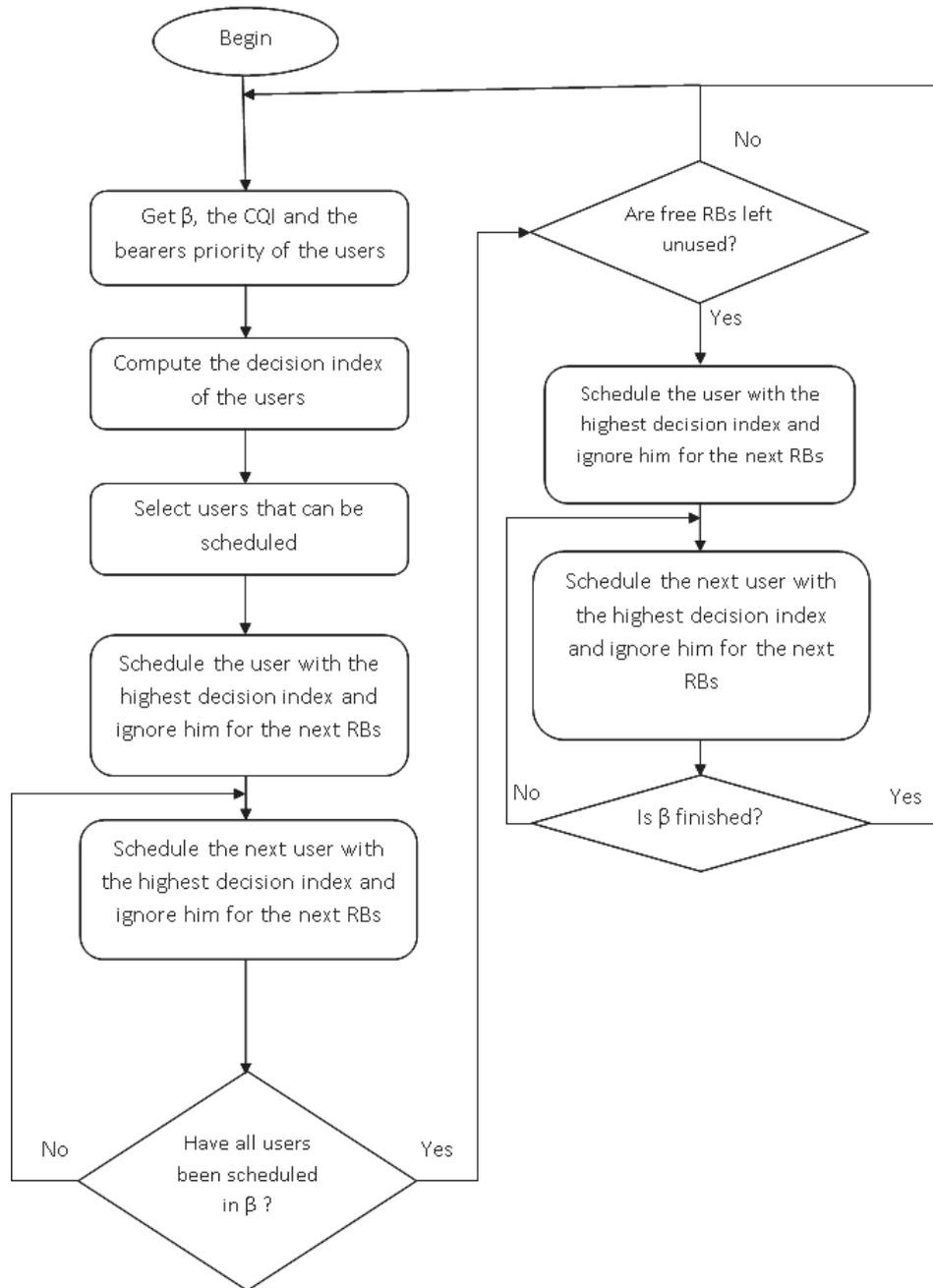


FIGURE 3.1: Flowchart of the proposed algorithm

model we assume also that the users share C resources blocks that eNB is responsible for the allocation. This allocation is done per-TTI basis and we have a small and limited amount of RBs due to the overbooking effect. The eNB uses a scheduling strategy to distribute the resources among the users. In our case, the scheduling strategy is based on the users channel conditions and the weight of their flow.

Due to the channel state variation and to adapt the transmission rate, we suppose that the eNB uses different MCS. For MCS modeling, we assume that the cell is divided

into k regions, each one corresponding to the use of a particular MCS. We assume $K = \{1, 2, 3, \dots, k\}$ and that the regions are concentric circles as shown in Figure 3.2.

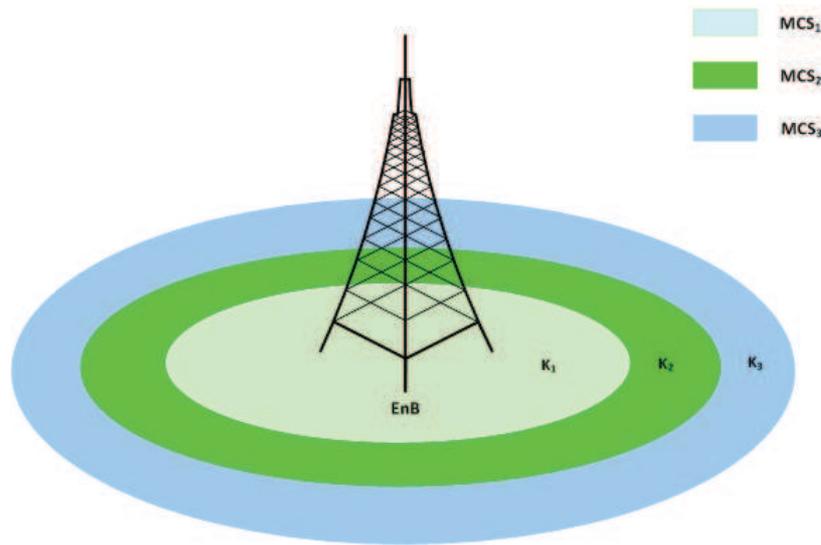


FIGURE 3.2: MCS distribution in the System

In this way at a given TTI, each user is present in a particular region k corresponding to the use of an associated MCS. Based on the k MCS and the m traffic classes, the scheduler allocates RBs using the configuration $Z_{k,m}$ of the users.

$Z_{k,m}$: User(s) present in region k and generating flow belonging to traffic class m .

In order to model the scheduling process, a Continuous Time Markov Chain (CTMC) of $m \cdot k$ dimensions with a state vector

$$\vec{n} = (n_{1,1}, \dots, n_{1,m}, n_{2,1}, \dots, n_{2,m}, \dots, n_{k,1}, \dots, n_{k,m}) \quad (3.6)$$

is used [42]. Each state describes the number of users of configuration $Z_{k,m}$ (user(s) present in region k and generating flow belonging to traffic class m) which is being scheduled.

We can easily see that our model becomes very complex (very large dimension and number of states) for any realistic value of k and m . For instance, LTE can have values of MCS going up to 15 and 9 different traffic classes. For these values, we obtain a CTMC of 135 dimensions. In the following, we use the minimal values of m and k to reduce the model resolution complexity.

Let us take a value of $m = 2$ and $k = 2$ which corresponds to a LTE downlink system with a cell divided into 2 concentric regions and where 2 different traffic classes can be used. As consequence, our system can be modeled using a 4D CTMC with a state

vector $\vec{n} = (n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})$ where each state describes the number $n_{k,m}$ of users of configuration $Z_{k,m}$ being scheduled. Since not all combinations of the number $n_{k,m}$ of users are allowed, let Ω represent the feasible states space. Ω contains all the states which verify the resources allocation feasible conditions. These feasible conditions depend on the maximum number of users of each $Z_{k,m}$ configuration which can be scheduled. This maximum number of users is closely related to the amount of resources that the scheduler allocates for each user of a particular configuration. Because of our scheduling strategy, the eNB takes into account the region where the UE is located along with the traffic class of its flow. For finding this amount of resources, we assume that traffic is modeled at connection level and each user uses only one traffic. We consider also that the system provides the necessary data rate for the different UEs applications. So, if a UE of configuration $Z_{k,m}$ is scheduled, the eNB will allocate the necessary $r_{k,m}$ resources to satisfy its data rate.

$r_{k,m}$ derives from the formula so that :

$$r_{k,m} = \frac{TTI \times \delta_m}{\beta_k} \quad (3.7)$$

Where δ_m is the data rate used to satisfy a UE generating flow belonging to traffic class m and β_k represents the number of bits given by one RB belonging to region k .

So if $S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})}$ is the state describing the possible $(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})$ combination, $S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \in \Omega$ if and only if:

$$(n_{1,1} \times r_{1,1} + n_{1,2} \times r_{1,2} + n_{2,1} \times r_{2,1} + n_{2,2} \times r_{2,2}) \leq C \quad (3.8)$$

1) Blocking States : The blocking states are set of states in which the scheduling of a new user would force a transition to a state $S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \notin \Omega$. Due to the fact that any user in the system belongs to a particular configuration $Z_{k,m}$, the definition of the blocking states is related to the configuration $Z_{k,m}$ of the user. Thus, we define the blocking states $S_{Z_{k,m}}^B$ of each $Z_{k,m}$ configuration:

$$\begin{aligned} S_{Z_{1,1}}^B &= \{S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \in \Omega \mid S_{(n_{1,1}+1, n_{1,2}, n_{2,1}, n_{2,2})} \notin \Omega\} \\ S_{Z_{1,2}}^B &= \{S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \in \Omega \mid S_{(n_{1,1}, n_{1,2}+1, n_{2,1}, n_{2,2})} \notin \Omega\} \\ S_{Z_{2,1}}^B &= \{S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \in \Omega \mid S_{(n_{1,1}, n_{1,2}, n_{2,1}+1, n_{2,2})} \notin \Omega\} \\ S_{Z_{2,2}}^B &= \{S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \in \Omega \mid S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}+1)} \notin \Omega\} \end{aligned} \quad (3.9)$$

This means that the scheduling of a UE belonging to the configuration $Z_{k,m}$ forces the state to move to a non feasible state.

We define also the blocking states S_m^B for the traffic classes. These blocking states correspond to set of states in which the scheduling of a new user generating flow belonging to a particular traffic class m (whichever its located region) would force a transition to a non feasible state.

$$\begin{aligned} S_1^B &= S_{Z_{1,1}}^B \cap S_{Z_{2,1}}^B \\ S_2^B &= S_{Z_{1,2}}^B \cap S_{Z_{2,2}}^B \end{aligned} \quad (3.10)$$

2) Markovian Model : As mentioned above, our system can be modeled using a 4D CTMC with a state vector $\vec{n} = (n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})$ where each state describes the number $n_{k,m}$ of users of configuration $Z_{k,m}$ being scheduled. The states space of the 4D CTMC are defined so that: $S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \in \Omega$ if and only if:

$$(n_{1,1} \times r_{1,1} + n_{1,2} \times r_{1,2} + n_{2,1} \times r_{2,1} + n_{2,2} \times r_{2,2}) \leq C$$

The chain can move to one of the four dimensions depending on the configuration of the new user which will be scheduled. Each state has several transitions and these transitions depend on the traffic model. Within this model, we assumed that traffic was modeled at connection level. So, transitions between feasible states happen due to call/session arrivals or due to call/session departures. We assume that the clients arrival can be modeled by a *Poisson process* with an arrival rate of $\lambda_{k,m}$ and service times are *exponentially distributed* with a rate of $\mu_{k,m}$. The state transitions diagram at a given state $S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})}$ is depicted in Figure 3.3

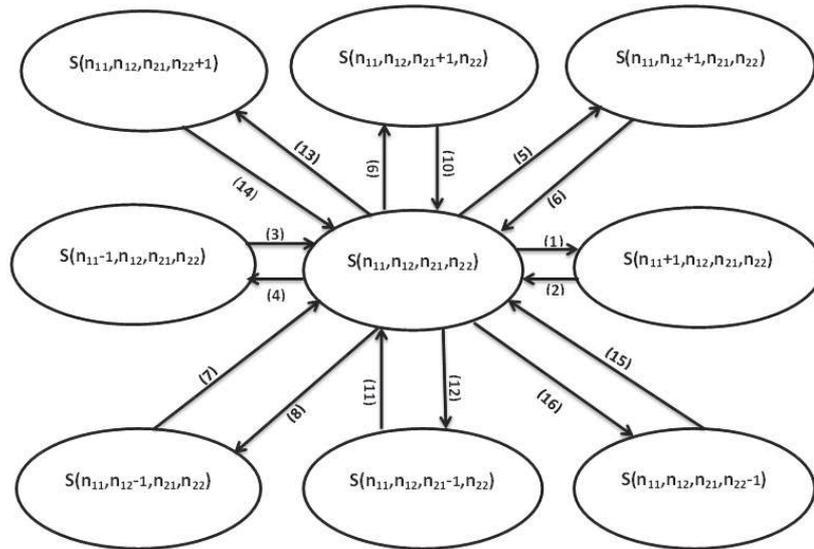


FIGURE 3.3: The Markov Chain

With

$$\begin{aligned}
(1) &= \lambda_{1,1} \\
(2) &= \mu_{1,1} \times (n_{1,1} + 1) \\
(3) &= \lambda_{1,1} \\
(4) &= \mu_{1,1} \times (n_{1,1}) \\
(5) &= \lambda_{1,2} \\
(6) &= \mu_{1,2} \times (n_{1,2} + 1) \\
(7) &= \lambda_{1,2} \\
(8) &= \mu_{1,2} \times (n_{1,2}) \\
(9) &= \lambda_{2,1} \\
(10) &= \mu_{2,1} \times (n_{2,1} + 1) \\
(11) &= \lambda_{2,1} \\
(12) &= \mu_{2,1} \times (n_{2,1}) \\
(13) &= \lambda_{2,2} \\
(14) &= \mu_{2,2} \times (n_{2,2} + 1) \\
(15) &= \lambda_{2,2} \\
(16) &= \mu_{2,2} \times (n_{2,2})
\end{aligned} \tag{3.11}$$

Due to possible non feasible states, we define function $\phi_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})}$ which guarantees that only feasible states are taken into account:

$$\phi_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} = \begin{cases} 1 & \text{if } S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

We can now deduce the balance equation for state $S_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} \in \Omega$ as :

$$\begin{aligned}
&P_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2})} [(1) \cdot \phi_{(n_{1,1}+1, n_{1,2}, n_{2,1}, n_{2,2})} + (5) \cdot \phi_{(n_{1,1}, n_{1,2}+1, n_{2,1}, n_{2,2})} + \\
&\quad (9) \cdot \phi_{(n_{1,1}, n_{1,2}, n_{2,1}+1, n_{2,2})} + (13) \cdot \phi_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}+1)} + \\
&\quad (4) \cdot \phi_{(n_{1,1}-1, n_{1,2}, n_{2,1}, n_{2,2})} + (8) \cdot \phi_{(n_{1,1}, n_{1,2}-1, n_{2,1}, n_{2,2})} + \\
&\quad (12) \cdot \phi_{(n_{1,1}, n_{1,2}, n_{2,1}-1, n_{2,2})} + (16) \cdot \phi_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}-1)}] = \\
&\quad (2) \cdot \phi_{(n_{1,1}+1, n_{1,2}, n_{2,1}, n_{2,2})} \cdot P_{(n_{1,1}+1, n_{1,2}, n_{2,1}, n_{2,2})} + \\
&\quad (6) \cdot \phi_{(n_{1,1}, n_{1,2}+1, n_{2,1}, n_{2,2})} \cdot P_{(n_{1,1}, n_{1,2}+1, n_{2,1}, n_{2,2})} + \\
&\quad (10) \cdot \phi_{(n_{1,1}, n_{1,2}, n_{2,1}+1, n_{2,2})} \cdot P_{(n_{1,1}, n_{1,2}, n_{2,1}+1, n_{2,2})} + \\
&\quad (14) \cdot \phi_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}+1)} \cdot P_{(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}+1)} + \\
&\quad (3) \cdot \phi_{(n_{1,1}-1, n_{1,2}, n_{2,1}, n_{2,2})} \cdot P_{(n_{1,1}-1, n_{1,2}, n_{2,1}, n_{2,2})}
\end{aligned} \tag{3.12}$$

By applying the constraint

$$\sum_{S_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \in \Omega} P_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} = 1 \quad (3.13)$$

we can obtain the steady states probabilities and compute the rewards of the system.

Making use of the blocking states previously defined, we compute the blocking probabilities of each configuration and of each traffic class.

$$\begin{aligned} P_{Z_{1,1}}^B &= \sum_{S_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \in S_{Z_{1,1}}^B} P_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \\ P_{Z_{1,2}}^B &= \sum_{S_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \in S_{Z_{1,2}}^B} P_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \\ P_{Z_{2,1}}^B &= \sum_{S_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \in S_{Z_{2,1}}^B} P_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \\ P_{Z_{2,2}}^B &= \sum_{S_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \in S_{Z_{2,2}}^B} P_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \\ P_1^B &= \sum_{S_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \in S_1^B} P_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \\ P_2^B &= \sum_{S_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \in S_2^B} P_{(n_{1,1},n_{1,2},n_{2,1},n_{2,2})} \end{aligned} \quad (3.14)$$

In order to determine the blocking probabilities, the steady-state probabilities are needed. For that purpose, we must solve equation 3.12 for all feasible states belonging to Ω . We use the Matlab solver to derive the steady-state probabilities and then use the same software to compute the rewards.

The analysis of the system rewards will allow to know whether or not the decision index defined in our strategy really impacts the resources allocation process. In order to achieve it, we use the minimum specified bandwidth for LTE (6 RBs) and evaluate the system under various traffic load $T_l = \lambda/\mu$, from 0.2 up to 1. To show traffic class differentiation, 64 kbps and 96 kbps are taken to specify the data rate used to satisfy traffic class 1 and traffic class 2, respectively. Finally, 128 bits/RB for region 1 and 64 bits/RB for region 2 are used in order to shape the region profile (number of bits given by a one RB belonging to a particular region).

Figure 3.4 shows the Blocking Probabilities of traffic classes for several traffic loads. We can observe that the blocking probabilities of each traffic class increase with the traffic load. In addition, we see that for each traffic load, the blocking probability of traffic class 1 is lower than the one of traffic class 2. At the beginning, there is not a great difference between the blocking probabilities but whenever the traffic load increases, the difference becomes larger. As example for $T_l = 1$, hardly 13% of the traffic class 1 calls are blocked and more than 18% of traffic class 2 calls are rejected. It means that more users generating flow belonging to class 1 are scheduled. This is normal because the bearer priority of the users flow is used by our scheduler for taking decision. Traffic class 1 which has more priority than traffic class 2 is ensured to have less rejected calls compared to traffic class 2.

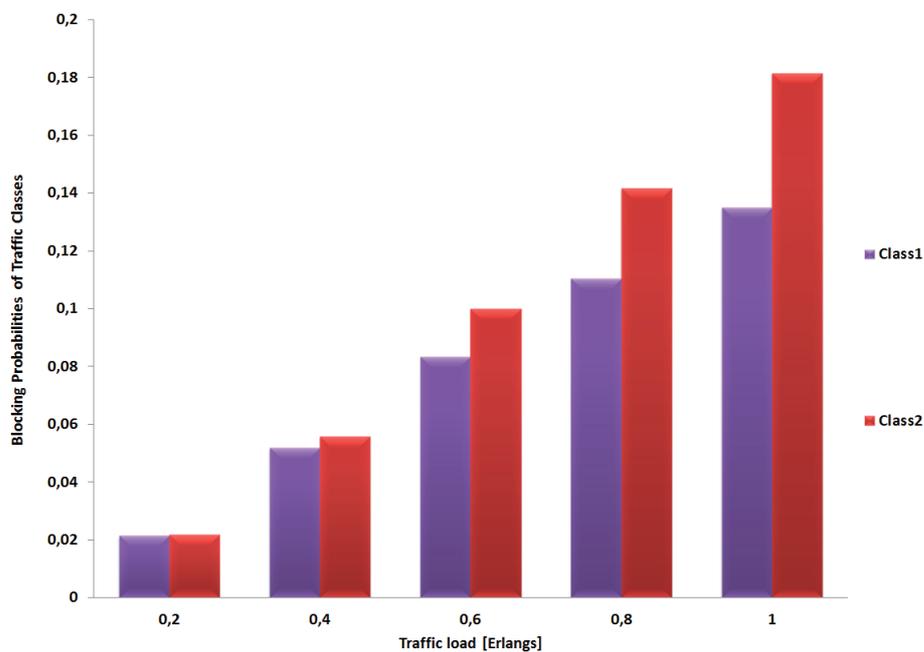


FIGURE 3.4: Blocking Probabilities of traffic classes

Besides the bearer priority, our scheduler uses also the channel condition for allocating resources. Thus, the users having the best channel conditions and the highest bearer priority have the best chance of being scheduled. This is confirmed in Figure 3.5 which shows the Blocking Probabilities that users of the different configurations undergo for the same traffic load. As expected, the users of configuration $Z_{1,1}$ are the most scheduled, whereas those of configuration $Z_{2,2}$ are the worst. Indeed, the users of configuration $Z_{1,1}$ experience the best channel conditions (region 1) and generate flow belonging to the highest bearer priority (Traffic class 1). But, those of configuration $Z_{2,2}$ are present in region 2 and have flow of traffic class 2.

These results show that our scheduler acts according to its design goal which is to distribute resources based on channel conditions and a QoS parameter.

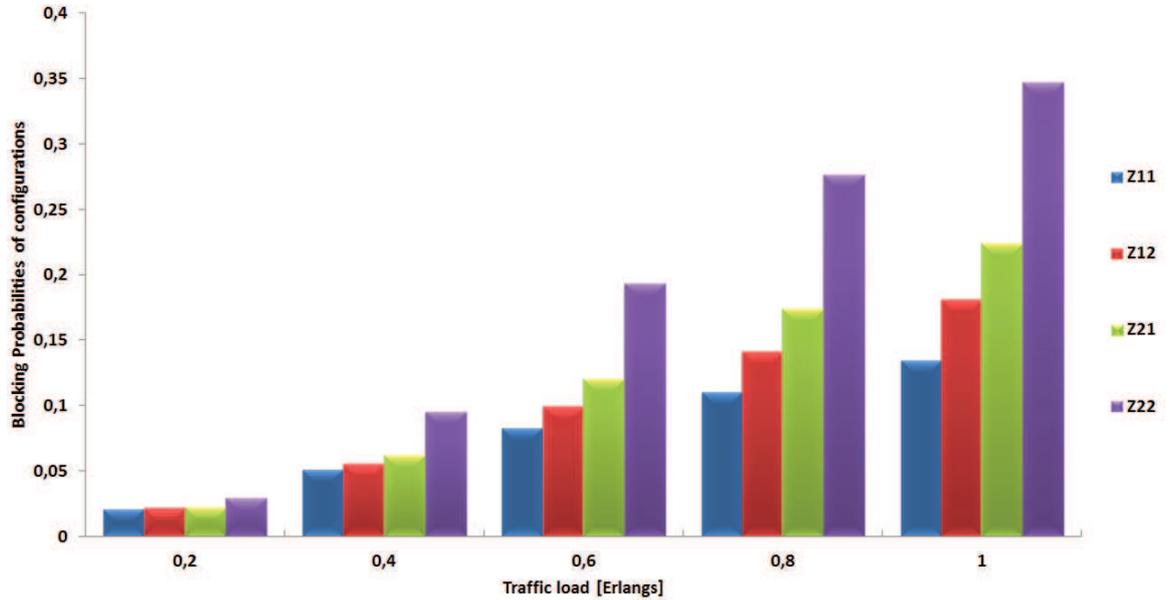


FIGURE 3.5: Blocking Probabilities of the different configurations

In the following subsection we will study the effectiveness of our strategy by assessing its performance through realistic scenarios.

3.2.4 Performance analysis

The performance of our scheme have been evaluated through simulations. NS-3 simulator ¹ has been used as simulation tools. It is a discrete-event and open-source network simulator, written in C++. It contains a LTE model which supports the evaluation of different aspects such as radio resources management, QoS-aware packet scheduling, inter-cell interference coordination and dynamic spectrum access. The simulations are used to compare our algorithm to the round robin method and the propotional fair scheme. In order to perform that, we use a minimum specified bandwidth for LTE (the 1.4 MHz band) where 6 RBs are available for resources allocation. To shape overbooking scenario, a number of users, from 8 up to 12 is considered (which is more users than available RBs). Localized type is used as RBs mapping type [10]. A single cell and SISO (Single Input Single Output) configuration with users using pedestrian mobility model (speed equal to 3 km/h) are also used. The simulation is done for various time periods (10,000 - 20,000 - 30,000 TTIs). These parameters are sufficient to evaluate the

¹NSNAM, <http://www.nsnam.org>

performance of our method as mentioned in [39] and [43]. The distribution of users positions and flows characteristics is done so that the first terminal is the closest user to the base station and has the highest flow priority and the last one has the worst channel conditions and the lowest flow priority. Within the simulation, each user receives one flow. The flows used are those defined by 3GPP in [14]. The set is composed of a conversational voice flow, a live streaming flow, a real time gaming flow, a buffered streaming flow, a ftp flow, an interactive gaming flow, a http flow and a progressive video flow. A summary of the flows distribution among the users is presented in Table 3.1. We measure the number of RBs and throughput achieved by the first terminal. We evaluate also the overall cell throughput and quantify the system fairness with Jains fairness index [44]. This configuration will help to study for the three algorithms, the impact of the overbooking scenarios on the system performance and on the performance of flows having the best parameters in terms of channel conditions and priority.

TABLE 3.1: Summary of the flows distribution

Terminal	Type of flows
User 1	Conversational voice
User 2	Conversational voice
User 3	Real time gaming
User 4	Live streaming
User 5	Buffered streaming
User 6	Ftp
User 7	Interactive gaming
User 8	Http
User 9	Progressive video
User 10	Interactive gaming
User 11	Progressive video
User 12	Buffered streaming

Figure 3.6 and Figure 3.7 show for all the schemes, the number of RBs and throughput achieved by user 1 during a simulation time of 10,000 TTIs. We want to point out that no performance changes have been observed after increasing the simulation time. We can observe that our algorithm outperforms the other algorithms and improves the performance of this user. In addition, the good performance of our algorithm are kept until reaching a number of 12 users. The overbooking effect is more experienced at RR level, because no parameter is used for allocation and users are scheduled sequentially.

By increasing the traffic load, the new users cannot have resources without decreasing the number of RBs of the old ones. It prevents some users from maximizing acquisition of resources. Regarding PF, it provides equal performance with our method at the beginning but after, the performance decrease whenever the traffic load increases. The reason is that PF does not take into account the nature of the flow to schedule but uses only information about channel conditions and fairness. In this way when the traffic load increases, user 1 has no guarantee to have more resources. With our algorithm, it is different because after serving all users, it schedules again certain UEs that are those which have the best decision index. User 1, whose flow has the highest bearer priority, always experiments the best decision index. So its flow is ensured to be scheduled more than the other flows. This is why user 1 keeps the same and high number of RBs until a twelfth user is added. At this level its performance decrease because after all users being served in the time window β , no free RBs are available. Indeed, in this case $\gamma = 0$.

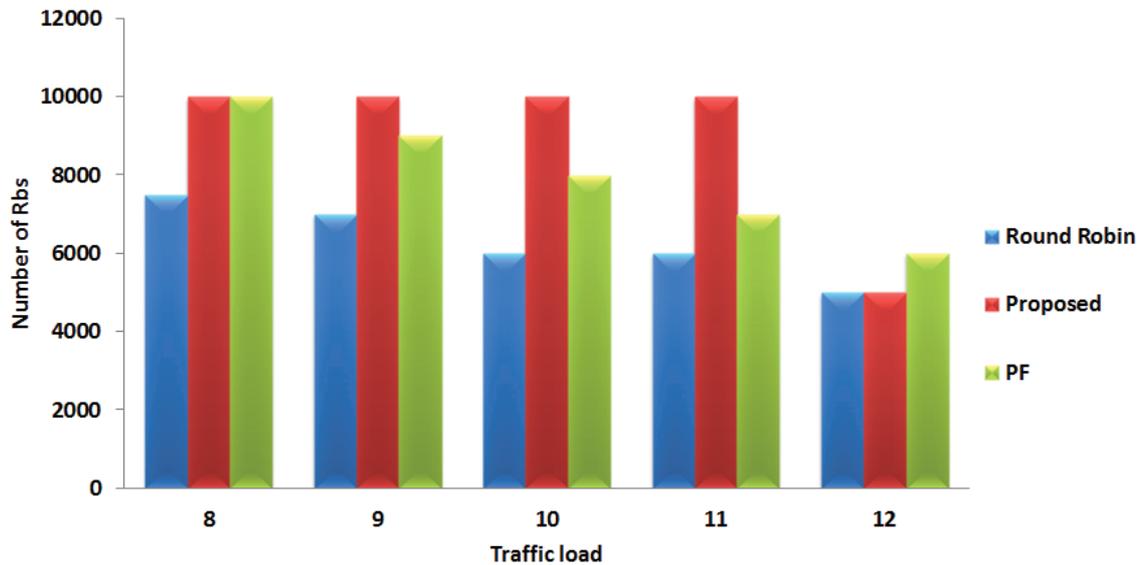


FIGURE 3.6: Best user's number of RBs

Figure 3.8 shows the total throughput achieved in the cell. Like previously, our method outperforms RR and PF algorithms. This is normal because in our scheme, we give priority to users with best channel conditions by scheduling them more than the others. The decreasing of the performance is due to the increasing of the overbooking degree. By increasing the traffic load, the number of privileged users (users with better throughput that are scheduled more than the others in β) is reduced. Additional simulations done at large scale (high traffic load) showed that our scheme kept providing better performance than the others until $\gamma = 0$. However in terms of fairness, the situation reverses as Figure 3.9 shows. We notice that the best fairness is achieved with the PF scheduler, which acts in accordance with its design goal. Our method achieves also a

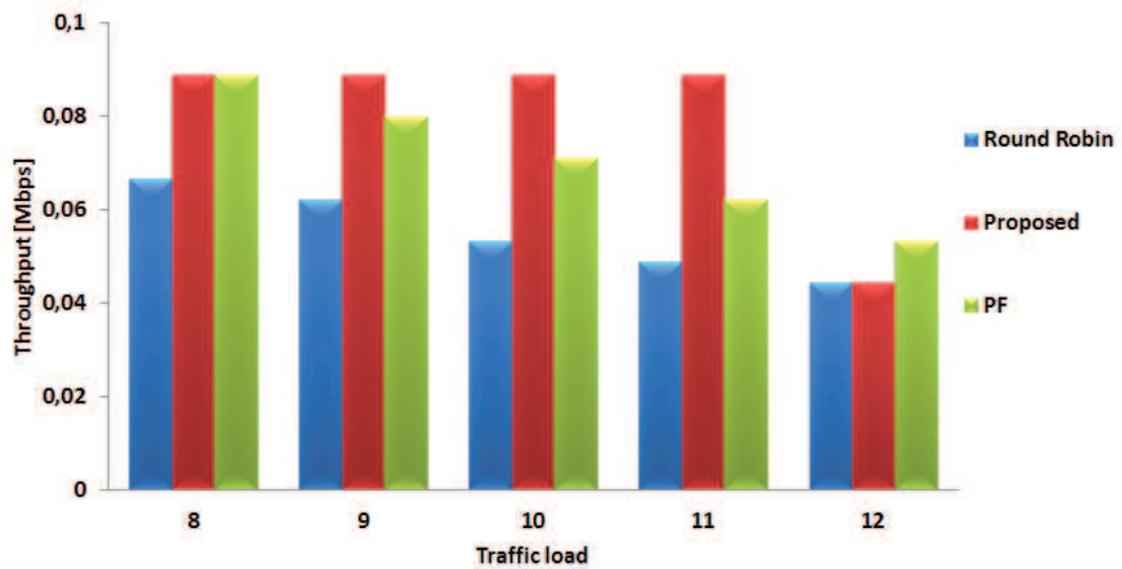


FIGURE 3.7: Best user's throughput

good fairness but never outperforms the PF, because there are significant differences between the users throughput. But as we increase the traffic load, the differences between the number of the users RBs are reduced and the fairness index increases. This is due to the goal of our method which does a compromise between fairness, overall system capacity and QoS requirements.

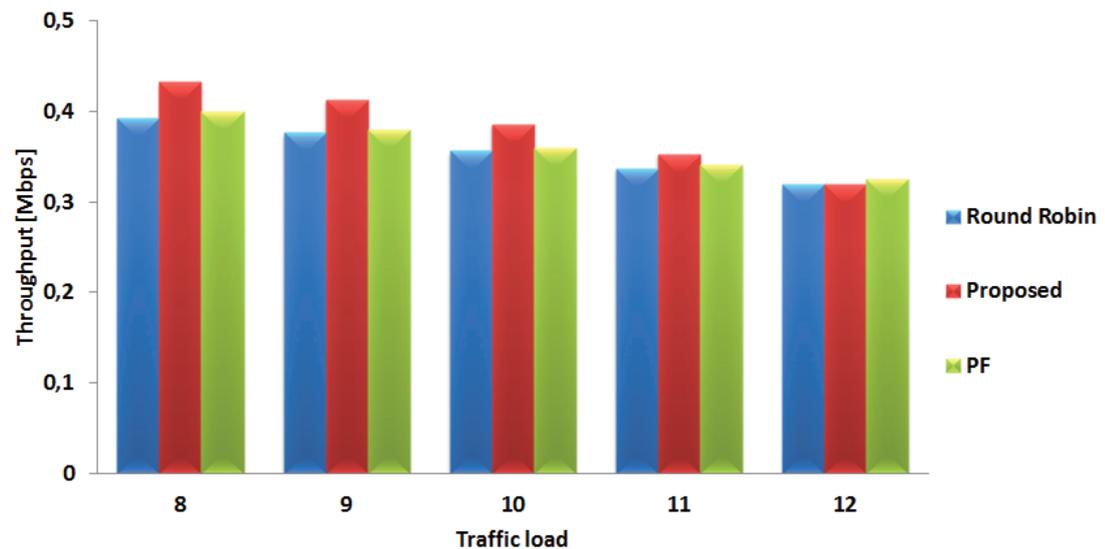


FIGURE 3.8: Overall cell throughput

All these results show that our scheme can be used in overbooking scenarios to improve the system throughput and give at the same time an acceptable level of fairness.

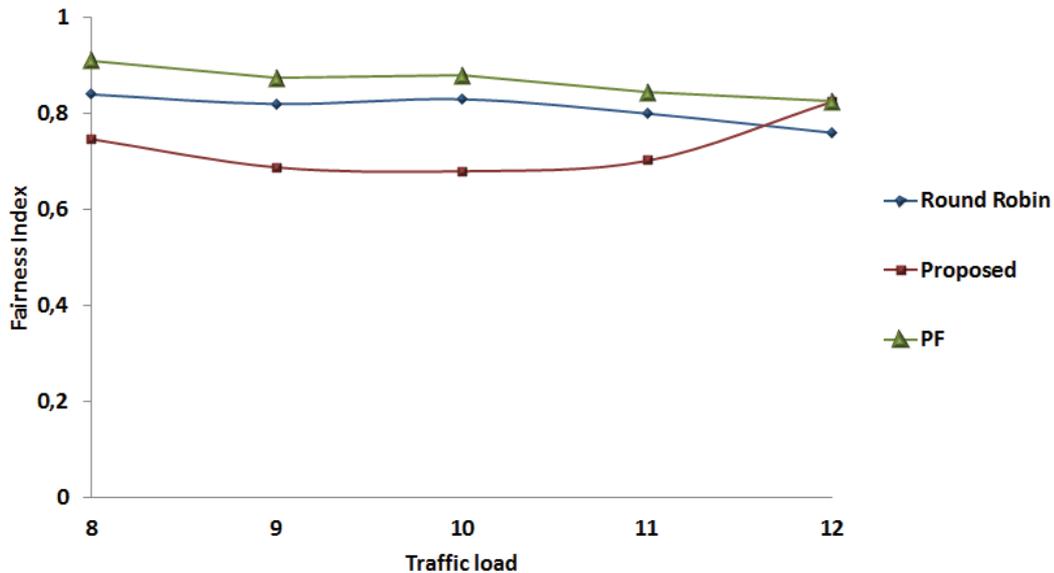


FIGURE 3.9: System fairness index

In this part, our work focused on improving resources allocation in case of overbooking scenarios. The proposed scheme allocates resources using the users channel conditions and the bearer priority of their flows. The key design aspects were first detailed and then we compared its performance with the RR and PF algorithms. The RR scheduling does not take into account any parameter for resources allocation, but rather schedules all users sequentially. It prevents users from throughput maximization. PF scheme deals with spectral efficiency and fairness but never considers any QoS parameter, thereby giving no guarantee to flows with high priority. The proposed method is more advanced and considers a QoS parameter for serving resources in overbooking scenarios. It uses a metric based on CQI reporting and bearer priority for taking decision. Fairness is enabled by using a time window and spectral efficiency is improved by scheduling twice in this time window the users with the best metric. Simulations have confirmed that our method allows an acceptable level of fairness while improving the overall cell throughput.

Although its performance in the overbooking cases, the scheme is not a complete QoS-aware strategy. Indeed the QoS parameter in the allocation decision is just a parameter used to give priority to users having strict QoS requirements and does not contain any information about the QoS requirements of each flow. Thus, our scheme could not be able to guarantee the meeting of the QoS requirements of the flows. It could not be also well-adapted to RT multimedia flows which require strict constraints on packet delay and packet loss. These applications are very important since they represent the most used applications in telecommunications. Therefore it is important and necessary to look into this issue and propose a scheduling strategy which aims at fulfilling the

QoS requirement of multimedia services. The following section will detail our proposed algorithm dealing with this issue.

3.3 An enhanced scheduling scheme for multimedia services

3.3.1 Context

LTE has become the most important radio access technology for mobile networks, providing significantly a ubiquitous broadband access. This breakthrough has to bring strong QoS support, especially for multimedia services which are the most used applications in telecommunication. Packet scheduling plays a major role to achieve this goal since it is responsible for resources allocation. Thus, an accurate scheduling scheme will be useful to multimedia flows by allowing to satisfy their QoS requirements.

Our first proposed scheme which has been designed for the overbooking cases, cannot be used for these types of services. Indeed as explained previously, although the fact that it uses a QoS parameter in the allocation decision, the used parameter does not contain any information about the QoS requirements of the flows. For instance the used QoS parameter does not take into account information about either the largest delay that packets of a RT flow can tolerate or the largest probability with which the delay requirement can be violated. However the aforementioned informations are very important in order to avoid packets to be dropped or to avoid packets to be delivered after the required delay. Therefore in presence of RT multimedia flows which require strict constraints on packet delay and packet loss, the proposed method will not be suitable.

For that reason, we propose an enhanced scheduling scheme which provides strict delay bounds and guarantees very low packet loss rate to multimedia services. This scheme is also made of two levels that work in series and is designed to improve the RT flows performance by providing the best QoS support.

3.3.2 Literature review

As mentioned above, resources allocation becomes more difficult in presence of multimedia flows since they require strict constraints on packet delay and packet loss. We have also seen in the scheduling strategies classification in Chapter 2 (see 2.6.3) that several researches have been done in recent years in order to overcome this challenge. Nonetheless to the best of our knowledge, a simple and reliable scheme able to allocate resources for satisfying very sharp delays bounds and providing very low packet losses, still has to come. We give below the description of these techniques and their limitations.

In [11] and [45], a simple scheme based on PF is proposed to maximize the total network throughput as well as assuring fairness among flows. It distributes resources based on the experienced channel quality and the past user throughput. Despite its efficiency, this method does not take into account any QoS parameter of the flows to schedule. So it becomes ineffective for multimedia services which have strict delay requirements. This limitation was overcome in [46] where an effective scheme for VoIP flows was designed. The key ideas consist in activating a VoIP priority mode and managing its adaptive duration. The VoIP priority mode allocates RBs first to VoIP flows in order to minimize VoIP packet delay and packet loss while the adaptive duration management helps to prevent the overall system performance degradation. Despite its performance, this scheme only focuses on VoIP flows and no other kinds of multimedia services are taken into account.

In [33] and [34], different scheduling strategies were proposed and focused on all type of multimedia services. The contributions have been very relevant to our work. They have been designed to increase the priority of RT flows as opposed to NRT flows. In other words, they have been proposed for the flows which have their head-of-line delay very close to the delay threshold. Information about the head-of-line delay is retrieved by the means of a function that changes according to the method used.

Finally in [4], authors used a different and innovative approach. They proposed a two-level framework called FLS (Frame Level Scheduler) that guarantees bounded delay to real-time flows. At the highest level, the scheduler computes the amount of data that each real-time flow should transmit in order to satisfy its delay constraint and at the lowest level, the scheduler allocates RBs to the delay constrained flows using a maximum throughput or proportional fair policy until the entire amount of data computed at the highest level has been transmitted. In spite of its effectiveness and originality, this scheme is not optimal in terms of reducing packet loss ratio. Indeed, since no QoS parameter is taken into account in the computation of the metric, some flows with pending data near the deadline could lose packets.

With all these schemes in the literature, scheduling multimedia services is still difficult taking into account the variations of the channel conditions as well as delay and packet loss sensitive services characteristics. Based on the method presented in [4], we propose an enhanced scheduling algorithm with the aim of improving the performance of multimedia traffic over the downlink of LTE networks. The following part describes how our method has been designed.

3.3.3 Key design aspects

The proposed scheme is made of two levels which act together [40]. The highest level is based on the one (FLS) that authors proposed in [4] and the lowest level is QoS-aware. We designed it to increase the performance of multimedia services by providing strict delay bounds and especially guaranteeing very low Packet Loss Rate (PLR). For that purpose, RT flows are firstly selected and given priority in the allocation of resources and once their scheduling finished, the remaining RBs are assigned to other types of services such as best effort. Figure 3.10 gives a general view of our scheduler.

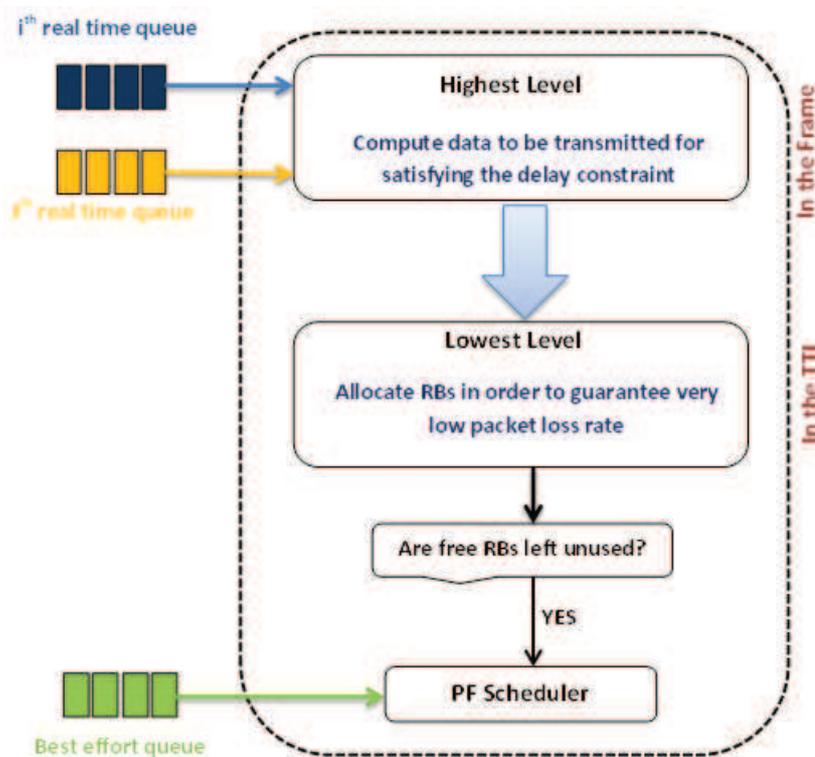


FIGURE 3.10: The enhanced two-level scheduler

At the highest level of the scheduler, a method defines frame by frame the amount of data to be transmitted by each RT source in order to satisfy the required delay and once this task accomplished, the lowest level assigns RBs every TTI to reduce packets loss, while considering the constraints of the upper level. In other words, the highest level works at the beginning of each frame by defining the amount of data that should be transmitted by each source. As for the lowest level scheduler, it is QoS aware and allocates RBs in each TTI in order to guarantee very low packet loss rate to the delay constrained flows. Once the multimedia sources have transmitted their data, the lowest level schedules the other flows in a proportional fair manner. The two parts below describe how the two levels of the scheduler have been designed.

3.3.3.1 The highest level of the scheduler

It has been designed using the linear discrete-time control theory [47]. With N active RT flows sharing the channel and having queues associated, the upper level evaluates the transmission needs of each queue every frame. In other words, at the beginning of each frame, a control law estimates the amount of data that the i -th RT source should transmit in this frame in order to meet its delay constraint. It is based on the assumption that:

$$\Delta_i = \rho_i \cdot T_f \quad (3.15)$$

where Δ_i is the upper bound of the delays of the i -th RT flow, ρ_i is the amount of data fragments of the i -th RT flow to send and T_f is the time interval used for sending one data fragment of the i -th RT flow (T_f is a LTE frame so lasts 10 ms).

The control law estimates for each data fragment to send in the k -th frame, the amount of data needed to satisfy the target delay. Thus, this amount can change according to the estimation of the control law. Indeed, it can be decided for instance to increase the amount of data that the RT source should transmit in the k -th frame and reduce this amount in the next frame.

3.3.3.2 The lowest level of the scheduler

It works every TTI and its role is to allocate RBs to the different types of flows (RT or not). According to the nature of the flow, its behaviour is different. Firstly, RT flows are selected and assigned RBs in accordance with the constraints of the upper level. It considers only those which have not yet transmitted their amount of data in the previous TTIs of the same frame. To satisfy these constraints and to guarantee very low packet loss rates, an exponential rule policy similar to the one mentioned in [11] is used. The metric of this policy is expressed by:

$$M_{LowestScheduler} = \text{Exp}\left(\frac{a_i \cdot D_{HOL,i}}{1 + \sqrt{\frac{1}{N} \sum_t D_{HOL,i}}}\right) \cdot \frac{R_{i,k}}{R_i} \quad (3.16)$$

with:

$$a_i = -\frac{\log \delta_i}{\tau_i} \quad (3.17)$$

where τ_i is the largest delay that packets of the i -th flow can tolerate and δ_i is the largest probability with which the delay requirement can be violated. $D_{HOL,i}$ represents the head of line packet delay of the i -th flow and $R_{i,k}$ and \bar{R}_i are the expected data rate for the i -th flow on the k -th RB and the past average rate achieved by the i -th flow respectively. N is the total number of active RT flows.

The presence of $R_{i,k}$ and \bar{R}_i helps to fulfill the constraints of the upper level and the presence of a_i and $D_{HOL,i}$ in the exponential pattern helps to reduce packets loss of flows with pending data near the deadline.

To explain in details the lowest level of our scheduler and to prove the truthfulness of our proposal, let us consider the queues of two RT flows which are going to be scheduled as shown in Figure 3.11. We assume that the packets in the flows queues arrive at different moment before being scheduled. We note $D_{HOL,1}$ and $D_{HOL,2}$ the head of line packet delay of flow #1 and flow #2 respectively.

We can see that:

$$D_{HOL,1} > D_{HOL,2} \quad (3.18)$$

We assume also that:

$$\delta_2 > \delta_1 \quad (3.19)$$

This means that flow #2 is allowed to have more packets which could violate the delay requirement without being discarded compared to flow #1 which have less. In other words, packets of flow #1 have the highest probability to be dropped if the delay threshold is violated. Considering that the two flows have the same delay threshold, we can conclude that:

$$a_1 > a_2 \quad (3.20)$$

Now let us see the resources allocation grids. Grid (a) and grid (b) show respectively RBs obtained by the two flows with a scheduler which does not consider any QoS parameter for taking decision and with our proposed scheme. It is clear that flow #1 which packets has the highest probability to be dropped, is guaranteed to have a very low packets loss rate in grid (b) than in grid (a).

At the end we notice that with the metric computed by our lowest level, besides indications on the rate that can be achieved by the delay constrained flows, QoS parameters are also taken into account for the allocation, thereby guaranteeing very low

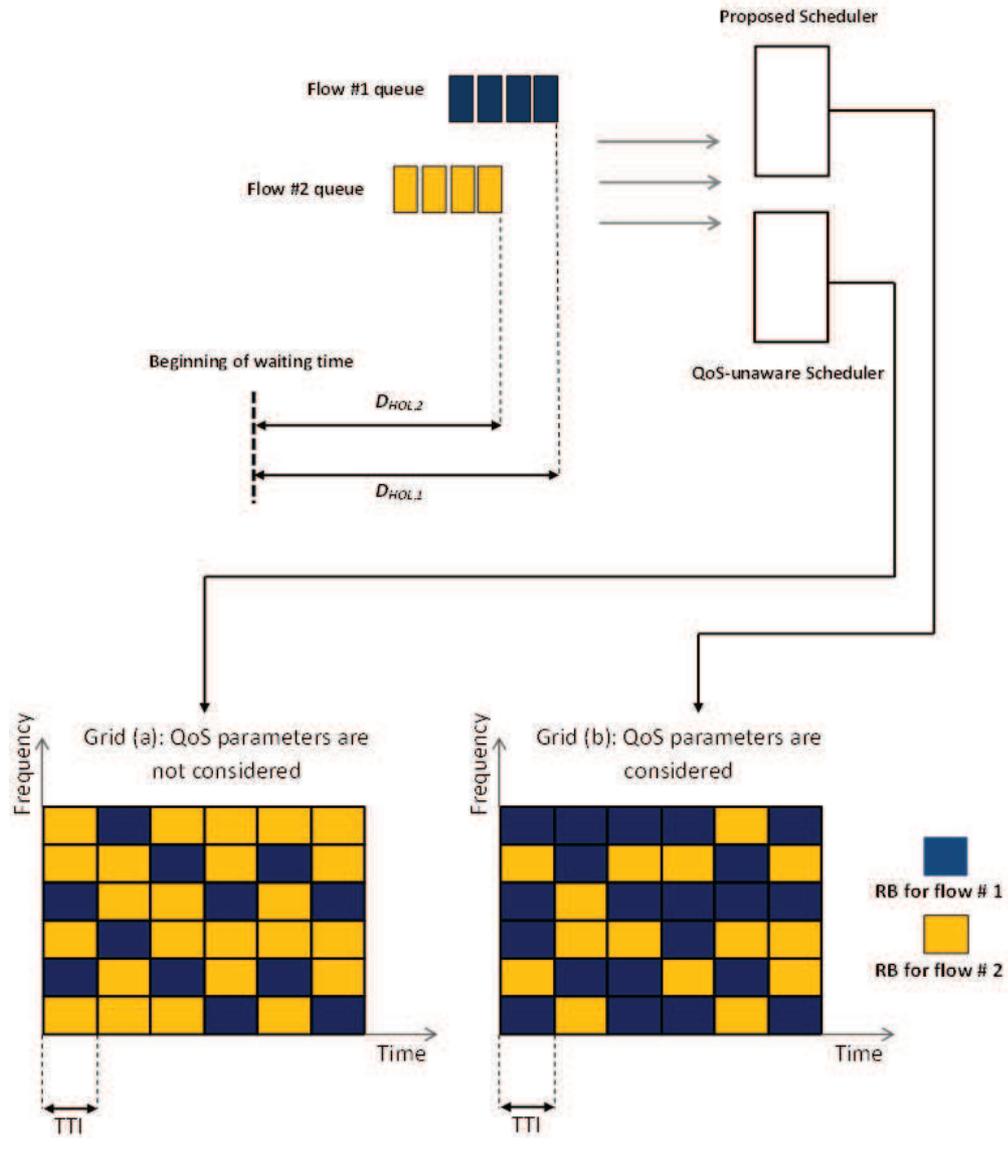


FIGURE 3.11: Example of scheduling process of two real time flows

packet loss rate. For RBs left free after the RT flows allocation, they are distributed to other types of flows in a proportional fair manner in order to maximize total network throughput and to guarantee fairness between them.

It is very interesting to design a new method that solves problems but it is better to study its effectiveness in realistic scenarios. This will be done in the following part.

3.3.4 Performance analysis

LTE-Sim [48] (see Appendix A) has been used to evaluate the performance of our method. It is a simulator that provides several aspects of LTE networks, including both

the LTE radio access network and the evolved packet core. It enables single and multi-cell environment, QoS management, mobility of users, handover procedures and frequency reuse techniques. The simulation results aim to compare our algorithm to the FLS framework [4], to a QoS-aware scheduling strategy (EXP-PF) and to the proportional fair scheme (PF). These results help to demonstrate the ability of our proposed strategy to provide very sharp delay bounds and guarantee very low packet loss ratio. Thus, we carry out simulations with different parameters which are shown in Table 3.2. The scenario consists in using a number of users in the range [10 40] that move along in a single cell. Each user has a velocity which does not exceed 3 km/h in order to stay in the low mobility context (pedestrian). Each user also receives three different downlink flows (1 video, 1 VoIP and 1 best effort). The CDF (Cumulative Distribution Function) of the packet delays of the RT flows, their average delay and their PLR (Packet Loss Ratio) are taken as key performance indicators. We consider these parameters because RT applications have no advantages in receiving expired packets and transmitting them after the deadline expiration [11]. This will be a waste of resources blocks. We evaluate also the total throughput achieved by these flows.

TABLE 3.2: Simulation parameters

Parameters	Value
<i>General parameters</i>	
Simulation duration	120 s
Users range	from 10 up to 40
Users speed (km/h)	3
Traffic models	H.264, G.729 and Infinite buffer
Mobility model type	Random direction
<i>LTE-related parameters</i>	
Transmission bandwidth (RBs)	25
MIMO	Off
Frame structure	FDD
Cell radius	1 km
Cell number	1
Target System	Downlink
Schedulers type	Proposed, FLS, EXP-PF and PF
Max delay	0.1 s

3.3.4.1 Traffic model

Trace-based applications are used as video flows. They send packets based on realistic video trace files (H.264) which are available on [49]. Voice flows are G.729 flows and they are modeled with an ON/OFF Markov chain, where the ON period is exponentially distributed with a mean value of 3 s and the OFF period has a truncated exponential probability density function with an upper limit of 6.9 s and an average value of 3 s [50]. During the ON period, the source sends 20 bytes sized packets every 20 ms (the source data rate is 8.4 kbps), while during the OFF period the rate is zero because the presence of a voice activity detector is assumed. Best effort flows are created by an infinite buffer application which is modeled like an ideal greedy source that always has packets to send. The part below presents the simulation results and our analyses.

3.3.4.2 Results and analysis

As shown in Figure 3.12 and Figure 3.13 for the two types of RT flows (voice and video), our method outperforms the other schemes by always reaching the lowest value for the PLR. For instance if we compare our strategy to the EXP-PF scheme, we can see that the PLR is reduced by 85,87% on the average for the video and by 37,61% on the average for the VoIP. We can also observe that the highest PLR reduction is obtained by comparing our scheme to the PF strategy. As for the FLS algorithm, despite its similarity with our scheme at the highest level, it was not able to provide the lowest packet loss rates. Indeed with our method, the PLR is reduced by 54,80% on the average for the video and by 17,22% on the average for the VoIP. This is due to the lowest level of our scheduler which gives priority to the RT flows having packets close to their deadline expiration. With this particularity, flows can expect to have a very low packet loss level. However, all the different approaches do not have the ability to reduce the PLR which increases with the growth of the traffic load.

The effectiveness of our scheme is also demonstrated with the CDF of the packet delays of video flows. We notice that no significant differences are observed with voice flows. Indeed, all the algorithms deliver the VoIP packets before the delay threshold (100 ms). However for video packets, it is different. In Figure 3.14 we can observe that only our algorithm and the other QoS-aware strategies (FLS and EXP-PF) always deliver the packets within the targeted deadline (100 ms). The proportional fair algorithm is unable to do it. We note also that our proposed scheme obtained the best performance. The difference is more obvious in the case of high traffic load. For instance when 40 users are used, more than 90% of the packets scheduled with our algorithm, has a delay lower than 50 ms, which is the highest rate compared to the other algorithms.

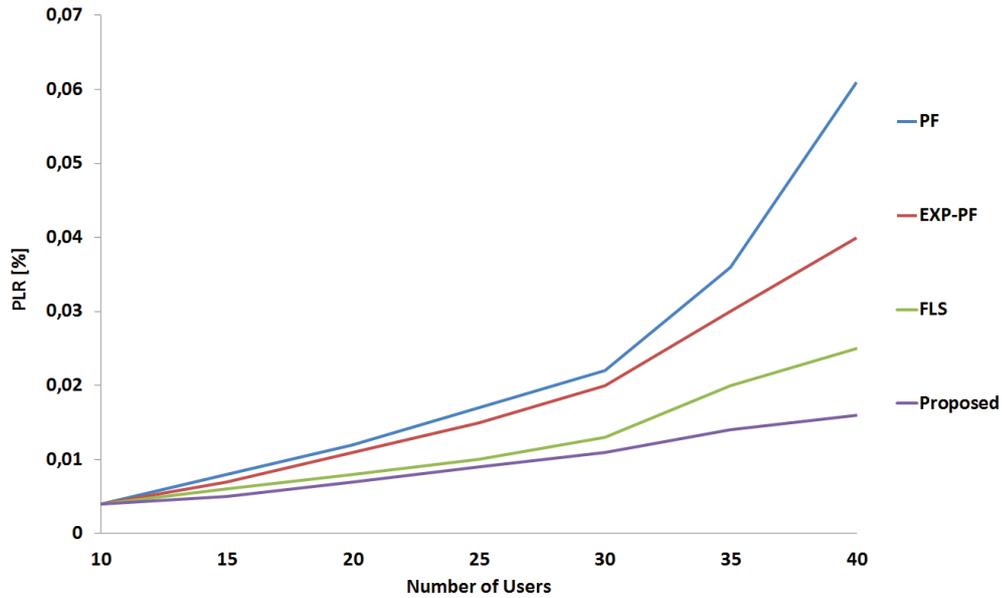


FIGURE 3.12: PLR of VoIP flows under several traffic loads

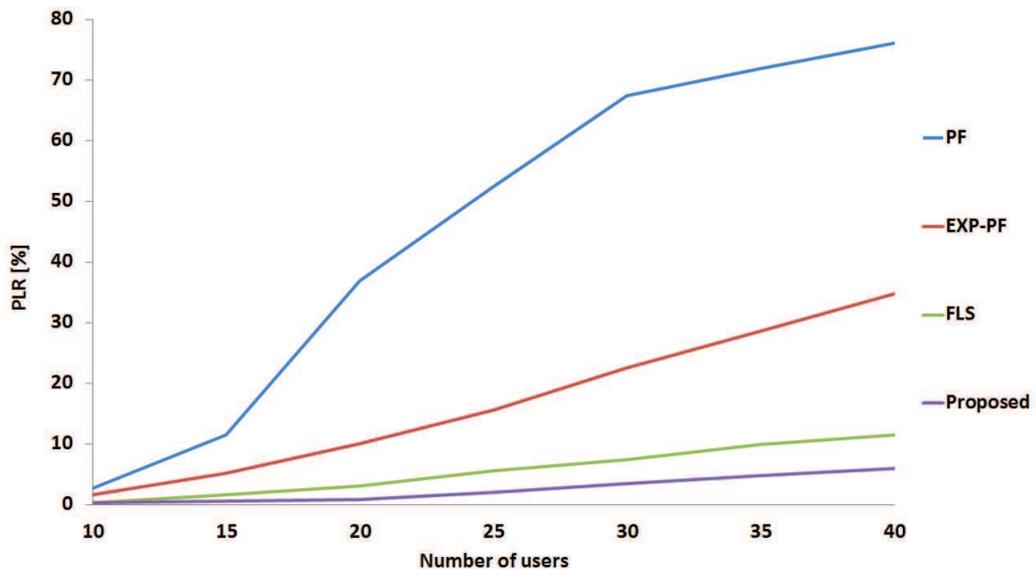


FIGURE 3.13: PLR of Video flows under several traffic loads

The efficiency of our strategy is also confirmed in Figure 3.15 which depicts the average delay of video flows scheduled with the FLS algorithm and with our proposed scheme. It can be seen that our method which is an enhancement of the FLS, outperforms it by reducing efficiently the average delay of these flows under several traffic loads. Indeed, with our method the delay is reduced by 13,33% on the average and the confidence interval of the delay reduction at 95% is [11,34% 18,62%]. It means that for any additional simulation which will be done outside the traffic load interval, it is sure at 95% that the

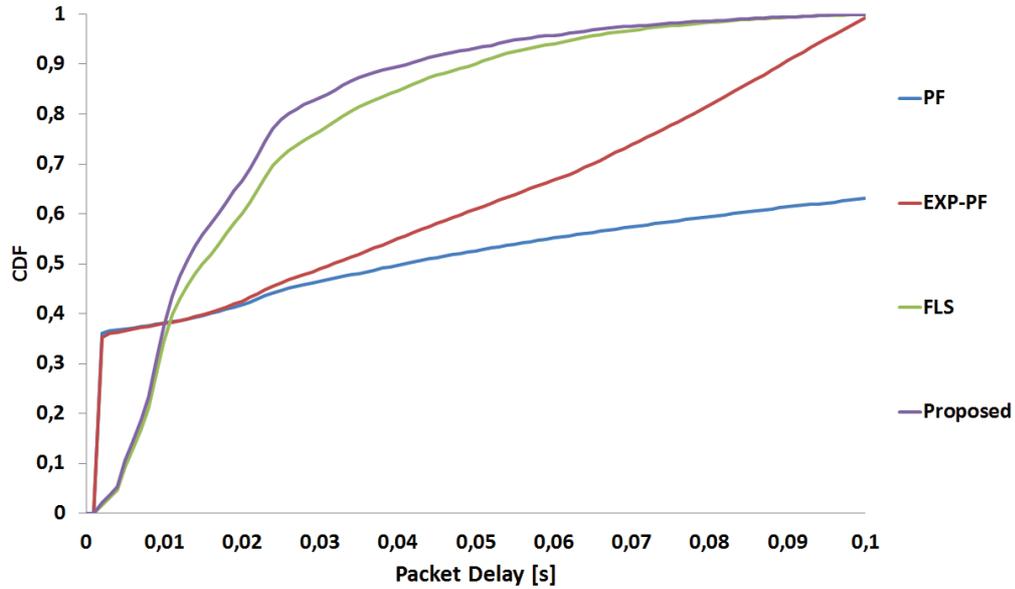


FIGURE 3.14: CDF of Video packet delays with 40 Users

delay reduction will fluctuate between 11,34% and 18,62% on the average.

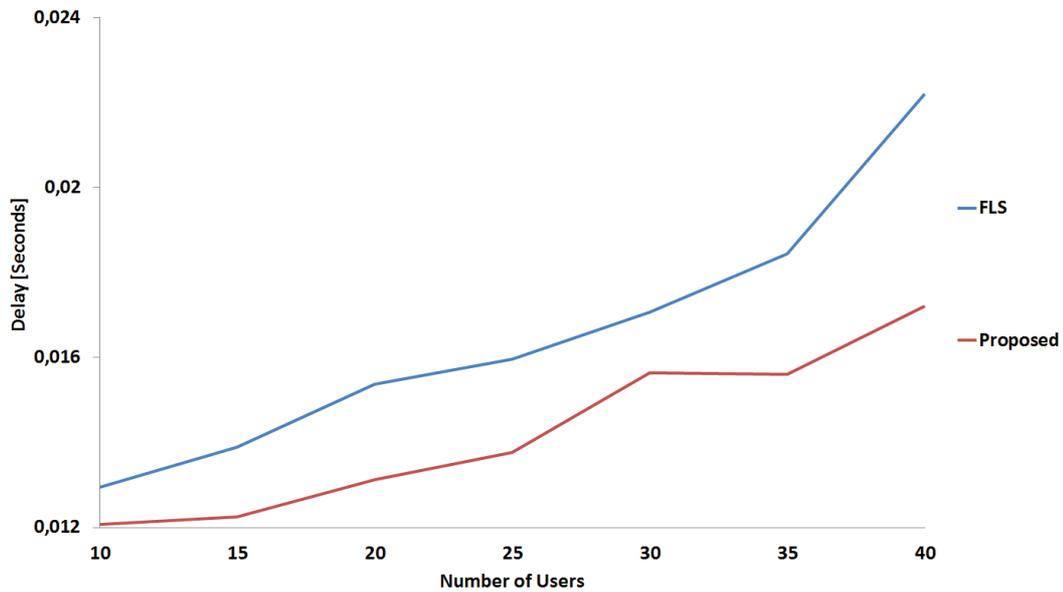


FIGURE 3.15: Average delay of Video flows under several traffic loads

By giving more resources to RT flows, we can observe in Figure 3.16 that the aggregate throughput of video flows computed with the QoS-aware strategies increases with the number of users. Especially for our method, the throughput obtained is the highest one among the other strategies. However for best effort flows, it is different. The aggregate throughput computed with our scheme decreases considerably with the traffic load as we see in Figure 3.17. In fact, our method is Pareto optimal [51]. That is to

say that it increases the resources allocated to RT applications at the cost of reducing those allocated for best effort flows. Furthermore, the aggregate throughput obtained for best effort flows with the PF is higher because PF does not take into account any QoS parameter for allocating resources, thereby providing no guarantees to multimedia flows and leaving thus, more resources for best effort flows.

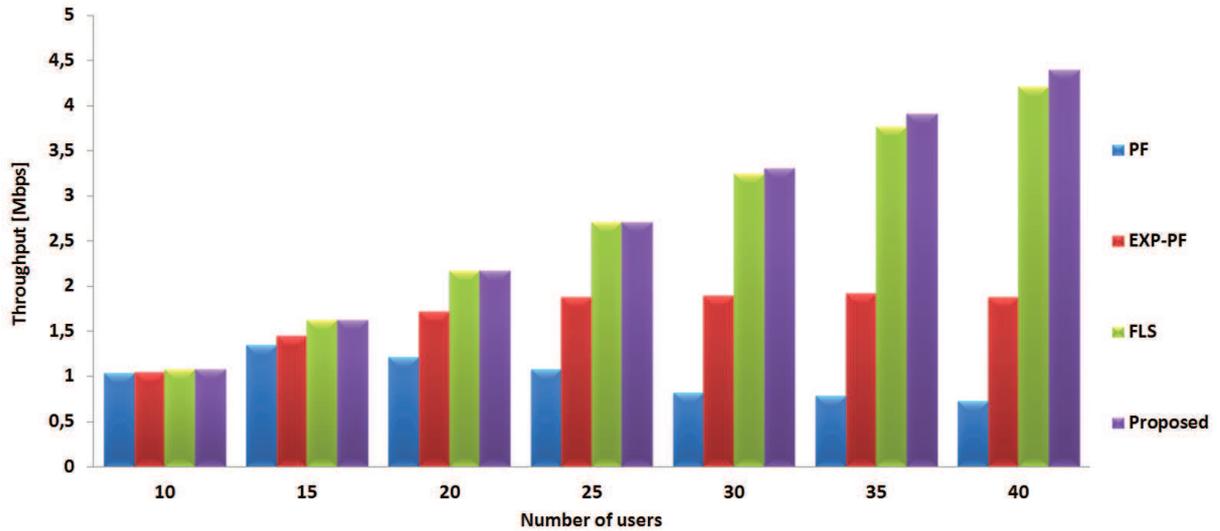


FIGURE 3.16: Aggregate throughput of Video flows under several traffic loads

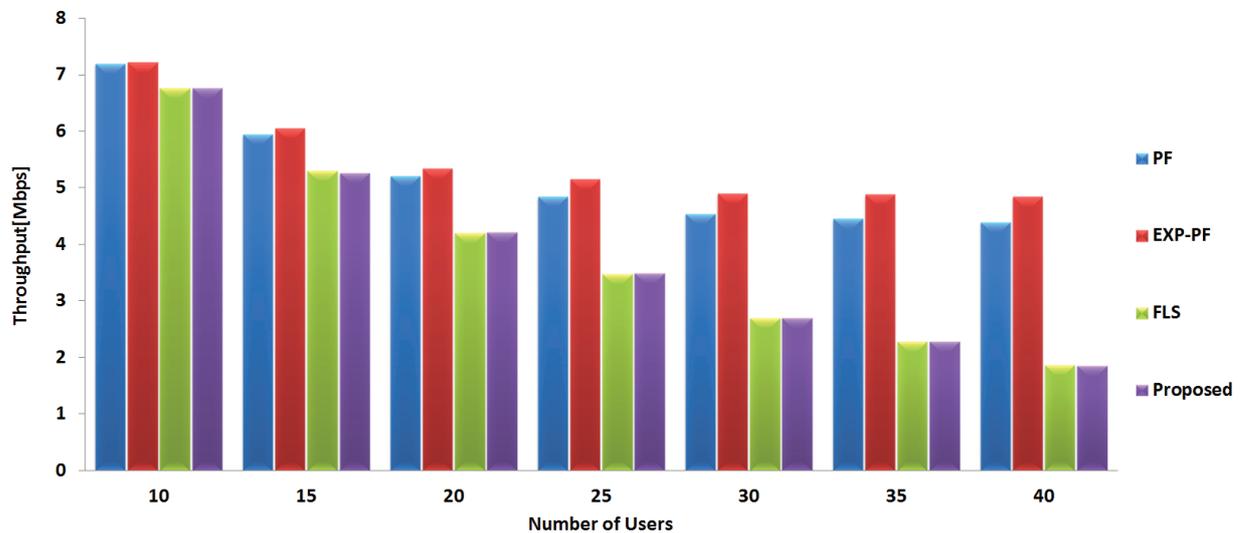


FIGURE 3.17: Aggregate throughput of Best effort flows under several traffic loads

In this part, we focused on scheduling issues for multimedia services in the downlink of LTE networks. We designed an efficient scheme made of two levels. At the highest level, the amount of data that each source should transmit in order to satisfy its delay constraint is computed and at the lowest level, the scheduler allocates RBs using an algorithm aiming at reducing the packets loss, while considering at the same time the constraint set by the highest level. Our proposed method has been evaluated and compared to other scheduling strategies. Simulation results have confirmed that our approach provided the best QoS support and increased multimedia services performances, especially for video flows. For instance, our method was always able to provide the lowest packet loss ratio despite the growth of the traffic load. In addition, the cumulative distribution function of the packet delays of video flows and their average delay showed that packets scheduled with our algorithm have the lowest delay.

3.4 Conclusion

Throughout this chapter, we addressed the problem of improving the scheduling of downlink communications in a context where the high mobility of the users is not taken into account. Generally it consists in designing strategies which satisfy the QoS requirements of all users by trying to reach, at the same time, an optimal trade-off between spectral efficiency and fairness. However the task becomes complex especially in overbooking scenarios where we have very limited resources for a great number of users, and also in presence of RT multimedia applications which require strict constraints on packet delay and packet loss.

To deal with these two issues, two innovative strategies were proposed in this chapter. Regarding the overbooking issue, a two-level scheduler was designed. This scheduler allocates resources using the channel conditions of the users and the bearer priority of their flows. It uses a metric based on CQI reporting and bearer priority for taking decision. Fairness is enabled by using a time window and spectral efficiency is improved by scheduling twice in this time window the users with the best metric. Simulations have confirmed that the method allows an acceptable level of fairness while improving the overall cell throughput in such scenarios.

To increase performance of RT multimedia flows which has strict constraints on packet delay and packet loss, we designed another scheme, since the previous proposed strategy was not a complete QoS-aware method and could not be able to guarantee the meeting of the requirements of such flows. The new proposed scheme is an enhancement of the FLS algorithm and made of two levels also. The highest level works at the beginning of each frame by defining the amount of data to be transmitted in order to

satisfy the required delay. As for the lowest level of the scheduler, it is QoS aware and allocates RBs in each TTI in order to guarantee very low packet loss rate to the delay constrained flows. Simulation results have confirmed that our approach provided the best QoS support compared to FLS algorithm, and increased also multimedia services performances, especially for video flows.

However as explained in Chapter 1, the designing of efficient resources allocation strategies should not be restricted only to downlink direction. It is necessary and important to look into the uplink direction. Indeed, the conception of effective scheduling algorithm in this direction is more complicated and represents a great challenge since the proposed methods should be compatible with the constraints set by this direction. In fact SC-FDMA which is used in uplink, requires the allocation to be done exclusively per users and the allocated RBs to be contiguous. Besides, the transmission process requires the eNB to be continuously informed about the amount of the users data to be transmitted, since it is not aware of this amount like in downlink. These constraints lead to the modification and the redesign of the scheduling strategies so that they can be efficient in the uplink direction.

In next chapter, we will present our novel strategy related to the scheduling process in the uplink direction.

Chapter 4

Resources allocation in the Uplink direction

This chapter presents our main contribution designed to improve uplink resources allocation of LTE networks in a context where the high mobility of the users is not taken into account. More precisely our solution is an innovative scheme which improves uplink resources allocation for videotelephony traffics and reduces the delay caused by dynamic scheduling. The key idea of this original method consists in scheduling videotelephony traffics using a semi-persistent strategy associated with a provisioning process. The performance of our proposed scheme have been evaluated by simulations in real LTE environments with LTE-Sim and simulations results demonstrated its effectiveness by showing that it optimized videotelephony traffics performance and provided the best QoS support compared to the dynamic scheduling.

4.1 Introduction

Nowadays, LTE has become the most important technology for mobile networks, providing significant and ubiquitous broadband access. It aims at bringing strong QoS support, especially for multimedia services which are the most used applications. For instance with LTE, RT flows will have a very low packet loss rate and all their packets will be delivered before the delay deadline. For this purpose, LTE introduces several new concepts and use various advanced techniques. The scheduling process is one of its main functions and plays a major role since it helps to efficiently allocate the RBs to reach the system performance and satisfy the QoS requirements of the services.

In the previous Chapter (Chapter 3), we focused on the downlink direction and proposed efficient strategies able to improve the scheduling process in order to provide the best QoS support to the RT flows. However as explained in this Chapter conclusion and following our thesis guideline, the designing of efficient resources allocation must be extended to uplink direction. Actually, in this direction the task is more complicated and represents a great challenge since the designed algorithms must be compatible with the constraints set by this direction. In addition, recent studies in [52] have shown that in the uplink of LTE networks, the handshake procedure consisting of a scheduling request message from the UE and a scheduling grant from the eNB requires twice a communication over the air interface and causes notable delay. This can be harmful for loss and latency-sensitive applications. Therefore in this Chapter, we address the problem of improving the scheduling of uplink communications in a context where the high mobility of the users is not taken into account.

We detail in this Chapter our innovative method which has been designed to deal with this issue. It is a novel scheme which improves uplink resources allocation for videotelephony traffics and reduces the delay caused by dynamic scheduling. Through the detailed presentation of uplink transmissions of videotelephony traffics, we highlight the lacks of the dynamic scheduling and suggest our new tailored proposition to meet the required needs.

4.2 A Semi-Persistent scheme for videotelephony traffics

4.2.1 Context

The recent years have witnessed a rising interest towards multimedia applications across the globe and has led to an important increase of mobile data usage and particularly a growing demand of video applications [53] [54].

To face this exponential growth, LTE has been proposed by the 3GPP and provides ubiquitous and significant broadband access. Today with LTE it is possible for the users to make natively one-to-one or one-to-many video calls, switch to video at any point during a call and drop video at any point to continue with just voice. This kind of video called videotelephony or conversational video is becoming more and more used and helps to increase the face-to-face collaboration and allow RT communications to take place at any location of the network. This is why it is predicted that two-thirds of the world's mobile data traffic will be video by the end of 2015 [7].

To bring strong QoS support with fast connectivity, high mobility and security to these video services, LTE introduced new features in the RAN such as OFDMA in the downlink, SC-FDMA in the uplink, the use of MIMO antenna schemes and a set of advanced MAC and physical functions. However, recent studies [52] have shown that in the uplink of LTE networks, the handshake procedure consisting of a scheduling request message from the UE and a scheduling grant from the eNB requires twice a communication over the air interface and causes notable delay. It was demonstrated that the RAN One Way Delay (RAN OWD) in the uplink of LTE networks was significantly higher compared to the one of the previous HSPA mobile networks. This could be harmful for loss and latency-sensitive applications such as videotelephony traffics.

For this purpose, we propose a new scheme which improves resources allocation for videotelephony traffics in this direction and reduces the delay caused by dynamic scheduling. The main contribution of this work is the design of a new scheduling protocol for handling videotelephony traffics in the uplink of LTE networks. The key idea consists in scheduling such traffics using a semi-persistent strategy with a provisioning process. The amount of resources to preallocate is estimated using an accurate traffic prediction model. It is just because such traffics in the network are characterized by variable sized packets with constant inter-arrival time. In addition the amount of their packets size can be predicted since these traffics consist of scenes in which people are talking with moderate motion in general and almost unchanged background [55]. Thus, it is possible to schedule these traffics in a semi-persistent manner and use traffic prediction models to preallocate resources.

Nethertheless, before detailing more the method we highlight the lacks of the dynamic scheduling through the description of the uplink transmissions of videotelephony traffics.

4.2.2 Uplink transmissions of videotelephony traffics

As explained in Chapter 2, the LTE uplink transmission scheme is based on SC-FDMA which has better PAPR properties compared to OFDMA. The PAPR characteristics are important for cost-effective design of the UE power amplifiers [19]. Scheduling of uplink resources is done by the eNB which allocates RBs or PRBs to the UEs every TTI of 1 ms in the time and frequency domain. Unlike downlink, resources allocation in uplink system is done per users. Besides, RBs to be assigned to the same UE must be contiguous. UEs ask for resources depending on their queue status. In the downlink, the eNB is obviously aware of the amount of data to allocate to the terminals but in the uplink it is different. The UEs have to inform the eNB about the amount of buffered

data to be transmitted and their priority. For that purpose, they send their BSR. UE's data is carried on the Physical Uplink Shared Channel (PUSCH) and UEs derive the uplink resources allocation after a handshake procedure consisting of a Scheduling Request (SR) message from the UE and a scheduling grant from the eNB. This process can cause notable delay and be harmful for latency-sensitive applications such as videotelephony. In the following we describe how resources allocation for videotelephony traffics is most commonly performed in LTE uplink (dynamic scheduling) and highlight the caused delay. The complete process is depicted in Figure 4.1.

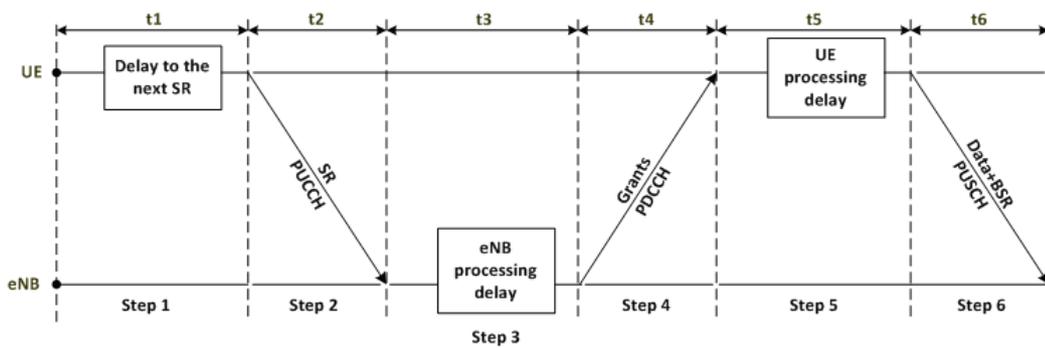


FIGURE 4.1: Uplink transmissions of videotelephony traffics with dynamic scheduling

- **Step 1: Delay to the next SR**

We assume here that a UE is becoming active and wants to send data related to its first video frame waiting in its buffer. As the UE has no resources on PUSCH, it has to send SR to the eNB. The SR procedure starts when a regular BSR is triggered but uplink radio resources to transmit the BSR are not available for the UE. During the SR procedure, the UE performs transmission of the SR over the PUCCH. The PUCCH resource for SR is allocated by the eNB in a periodic manner. The periodicity of the PUCCH resource allocated for SR is called SR periodicity. The SR periodicity impacts upon the delay for the UE to obtain uplink resources and is configured by the eNB. So depending on its SR periodicity configuration, the UE has to wait a certain time before sending the SR. We consider as t_1 the delay to the next SR opportunity.

- **Step 2: SR transmission**

UE transmits its SR to the eNB. We assume t_2 as the transmission time.

- **Step 3: eNB processing delay**

After receiving the SR, the eNB has to decode the SR and then generates the grants for the UE. We notice that for the first grant to generate, as the eNB has no information about data that the UE has to transmit, it gives minimal grants for the UE. This grant

will allow to send primarily the BSR along with a part of data waiting in the buffer. We assume t_3 as the eNB processing delay.

- **Step 4: Grants transmission**

The eNB transmits the uplink scheduling grants over the PDCCH. We assume t_4 as the transmission time.

- **Step 5: UE processing delay**

The grants will allow the UE to transmit data over the PUSCH, but first at all, it has to decode the grants sent over the PDCCH and after encode the data to transmit. We assume t_5 as the UE processing delay.

- **Step 6: Data transmission**

UE transmits the BSR along with a part of data waiting in the buffer to the eNB over the PUSCH. We assume t_6 as the transmission time.

Finally the total delay T before transmission is :

$$T = t_1 + t_2 + t_3 + t_4 + t_5 + t_6 \quad (4.1)$$

The value of the total delay is not fixed and may vary due to the variation of the delay of the different steps. For assessing this total delay, we will base on estimations which can be found in 3GPP documents, in scientific papers and industrial reports [1] [56] [57].

- **t_1 estimation**

As mentioned early, the delay to the next SR opportunity depends upon the SR periodicity configuration. In our analysis, we assume that the UE is configured with the basic SR periodicity of 5 ms. So, the average delay for sending the SR is 2.5 ms [57].

- **t_2 estimation**

t_2 is the transmission time for sending the SR. For LTE, this time is fixed and is equal to the TTI duration. $t_2 = 1$ ms

- **t_3 estimation**

t_3 is the eNB processing delay. This processing delay depends upon the time taken for decoding the SR and for generating the grants. The eNB generates the grants after taking its decision which is based on the use of the scheduling algorithm. We assume that the eNB takes 3 ms to accomplish this task [1] [57].

- **t_4 estimation**

t_4 has the same value as t_2 . In this way, $t_4 = 1$ ms

- **t_5 estimation**

t_5 is the UE processing delay. We assume that the UE takes at most also 3 ms to decode the grants sent over the PDCCH and encode the data to transmit [57].

- **t_6 estimation**

t_6 is the time used for transmitting the uplink data. We note that it is not sure that the UE will be able to transmit all enqueued packets within the single TTI. In general t_6 depends from several aspects which comprise the scheduling algorithm (how often the eNB allocates uplink resources to a UE), the channel quality (the number of bits that can be transmitted within each RB), the number of allocated resources in each TTI and the queue size. In this work we assume that $t_6 = 1$ ms [1] [57].

With this configuration, \mathbf{T} is approximately equal to 12 ms. We note that this value is based on assumptions and in reality the value of \mathbf{T} is generally higher than the assumed value [52]. We just give a view on how long can be the user plane latency with dynamic scheduling.

For the videotelephony traffic, it means that the first video frame in the UE buffer has to wait roughly this time (12 ms) before to be sent. Thus, the transmission of this first video frame can take a long while because of the latency caused by the handshake procedure. In addition we notice that it is not the entire amount of data of the first frame which will be sent at the first occasion because the first grant is minimal and is given primarily to send the BSR. So in the best case (resources are always available for UE after sending BSR and UE has a very good channel quality), at least two transmissions on PUSCH will be used for sending the entire amount of the first frame. But in reality when dynamic scheduling is activated, resources are not always guaranteed for a UE and it can take more PUSCH (more time) for sending the first frame. It depends principally on the scheduling strategy implemented in the eNB and another parameter which is the channel quality (the number of bits that can be transmitted within each RB). As consequence, the next frames, after being available in the UE buffer, could wait for a while and be delayed. This could increase the user plane latency and cause at the same time significant end-to-end delay. Our proposed strategy will help to improve resources allocation for videotelephony traffics and reduce the delay caused by dynamic scheduling. However before going to the heart of the proposed method, we present an overview of the existing methods and algorithms designed for solving uplink resources allocation problems.

4.2.3 Literature review

Scheduling strategies play a major role in LTE systems because a great performance gain can be achieved by properly allocating resources among users. This allocation has to take into account various aspects such as meeting the expected QoS, assuring fairness and maximizing the total network throughput. It becomes more difficult in presence of multimedia services such as videotelephony flows which require strict constraints on packet delay and packet loss. Several researches have been done in recent years in order to overcome these challenges. However, most of these works focused on the downlink direction. Nevertheless, researches have been conducted to design appropriate scheduling schemes in the uplink direction. In the literature the proposed schemes can be classified into three different groups which are opportunistic, delay-based and multiclass-based [23].

The first group is composed of opportunistic scheduling strategies. As seen in section 2.6 of Chapter 2, these strategies exploit the variant nature of the radio channel conditions to decide which time slot to transmit data for each UE. Several schemes based on such approach have been proposed in [58]. The PF algorithm considered as the reference and used in the downlink direction has been modified and adapted for the uplink direction. Indeed, the fact related to the contiguous resources block allocation required by SC-FDMA forces to modify this approach. In this way, two promising strategies, namely FME and RME have been proposed in [31] and [32] and are based on the PF utility function. FME algorithm allocates the RBs to a UE beginning from the RB where the UE has the highest metric value in the resources allocation matrix and expanding the assignment on the right and the left sides of this RB. The resources allocation of the UE is stopped whenever the algorithm finds another user having a higher metric on the considered RB. RME acts like FME scheme but with the particularity of performing a recursive search of the maximum metric. Both algorithms show good performance in terms of general throughput. However, performance in terms of packet delays and packet losses have not been tested.

The second family is composed of delay-based scheduling strategies. They are QoS-aware and were proposed for all type of multimedia services. These methods take their decisions by exploiting the packet delays and head of line values. Information about the packet delay is retrieved by the means of a function which changes according to the method used. Several mechanisms based on this strategy have been proposed in [59] and [60]. In [60] the proposed mechanism considers the packet delays as the main metric for taking decision and computes them in a different way. Usually packet delays are computed by taking into account queues waiting time, while this scheme computes the

packet delays by using estimations. The obtained results using these algorithms show a good performance in terms of QoS support.

The third family represents the multiclass-based scheduling strategies. With these schemes, the scheduler takes into account the UEs flows classes for performing the resources allocation. In other words, the RBs allocation is performed once the scheduler has checked the type of service of the flow. Using these strategies, UEs handling RT flows can be prioritized. As consequence, these schemes help to provide the best QoS support to RT flows but are not optimal in terms of fairness, since the NRT flows are neglected.

With all these promised scheduling schemes in the literature, managing videotelephony traffics in the uplink is still difficult taking into account the constraints required by the uplink direction (SR messages from the UE and scheduling grants from the eNB) and the variations of the channel conditions as well as the delay and packets loss sensitive services characteristics. Indeed, although the fact that an accurate scheduling method can be used in order to provide strong QoS support, this one could become ineffective if a notable delay already occurred. A solution would be to reduce this delay caused by the handshake procedure between the UE and the eNB by designing a new scheduling protocol. To the best of our knowledge, no previous works have been done for the videotelephony traffics in the uplink of LTE networks. Thus we propose in this chapter a new scheduling protocol which focuses on these types of traffics. The section below gives more details about the new scheme.

4.2.4 Key design aspects

In what follows, we explain the novel scheduling protocol strategy by firstly detailing how works the protocol and after presenting the prediction model used for the videotelephony traffics. The last part consists in highlighting the scheduling algorithm aspects.

4.2.4.1 Protocol

Our goal by designing this strategy is to improve the videotelephony traffics resources allocation by reducing the delay caused by the dynamic scheduling in LTE uplink. For that purpose, our scheme focuses mainly on the video frames which come after the first frame and aims at reducing considerably their waiting time in the buffer. All the process is presented in Figure 4.2.

- **Preliminary step**

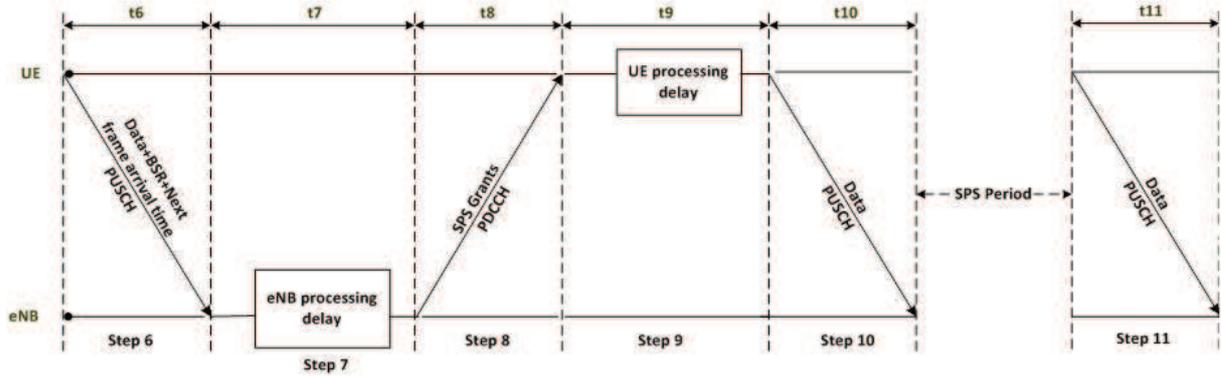


FIGURE 4.2: Uplink transmissions of videotelephony traffics with SPS-P

Due to the fact that Variable Bit Rate (VBR) videotelephony traffics in the network are characterized by variable sized packets with constant inter-arrival time, we choose a semi-persistent mode with provisioning to allocate resources to the frames which come after the first frame. With the semi-persistent mode, the UE will not request RBs for these frames to come and the provisioning mode will permit to define the amount of resources to reserve for each next frame using an accurate traffic prediction model. For the implementation of our strategy, we define a new scheduling mode called Semi-Persistent Scheduling with Provisioning (SPS-P). This mode has to be configured by the eNB (RRC connection setup message) once a videotelephony user is connected to the network and activated when required (See steps below). As mentioned early, the strategy focuses on the frames coming after the first frame. So it takes effect from the second frame. Due to few opportunities for the UE, the eNB will be in charge of forecasting the amount of resources to reserve. However, the success of this forecast has two main constraints which are time and data. Time because the eNB has to finish the forecasting and send the grants before the arrival of the second frame. Data acquisition will help to use the traffic prediction model to compute the bandwidth requirement of the next frames. This is why the UE will assist the eNB for the success of this task.

- **Process**

After decoding the grants sent over the PDCCH and encoding the data to transmit, UE must also estimate the waiting time of the first frame. Indeed, the UE knows how long the first frame is waiting in its buffer. In **step 6**, the UE transmits the BSR, the estimation of the waiting time of the first frame and a part of data waiting in the buffer to the eNB over the PUSCH. **Step 7** is the milestone in the success of our scheme. It helps to implement our strategy. Moreover we note that as the same time as this step is performed, eNB keeps sending dynamic scheduling grants until the entire transmission of the rest of data of the first frame.

To achieve **step 7**, the eNB has to predict the amount of resources to reserve for the next frames. In order to predict this amount of resources, the eNB uses a traffic prediction model to compute the bandwidth requirement of the next frames. However, the eNB cannot compute bandwidth requirements indefinitely. The prediction window is limited by the arrival of the next frame (the second frame). In other words, the number of next frames which will have preallocated resources is limited by the time available for the forecasting computations. The eNB can estimate this time using information given by the UE in **step 6**. Indeed, due to the fact that the frame inter-arrival time is constant and the transmission time in LTE is fixed, the UE can deduce this computation time.

If we assume T_f the information given by the UE in **step 6** (waiting time of the first frame), T_c the time available for computations, T_g the time used for generating the grants and F_t the frames inter-arrival time, we can deduce that :

$$T_c \leq F_t - T_f - T_g - t6 - t8 \quad (4.2)$$

After deducing this time, the eNB knows how long at most it could take for the next frames resources prediction. Based on this interval of time and using also the traffic prediction model, it computes the amount of resources to reserve for the next frames.

If we assume P_t the time used for predicting the amount of resources which will be allocated to one frame, the prediction window N_f (number of next frames which will have preallocated resources) derives from the formula:

$$N_f \in \mathbb{N}, N_f \leq \frac{T_c}{P_t} \quad (4.3)$$

Once the resources prediction of the next frames is done, the eNB is ready for activating SPS-P. It generates the grants associated and then sends them to the UE over the PDCCH in **step 8**. UE will use these recurred resources in accordance with the indications of the SPS-P parameters. The main parameters are the transmission period and the number of transmission that UE is allowed to perform.

The SPS-P transmission period corresponds obviously to the frames inter-arrival time and the number of transmission is limited by N_f . It means that resources will be reserved for the N_f next frames after the first frames. Two other parameters are defined to control the data transmission, namely the threshold S_t for characterizing an inaccurate transmission and the number N_i of consecutive inaccurate transmissions which is not allowed. They are used to prevent the waste of resources. Indeed, the UE releases the resources when a certain number of consecutive inaccurate transmissions is

carried out. An inaccurate transmission is a transmission where the ratio R_i between the amount of data to send and the amount of preallocated resources is lower than the threshold allowed (over allocation).

A transmission is considered as inaccurate if $R_i < S_t$ and the SPS-P mode will stop when reaching N_i consecutive inaccurate transmissions. In addition, to avoid pending data when the amount of preallocated resources is lower than the amount of data to send, the UE keeps sending BSR along with the PUSCH to help the eNB to dynamically schedule transmission of this pending data (under allocation).

4.2.4.2 Prediction model

The proposed scheduling protocol requires that the eNB forecasts the amount of resources to reserve for the next frames. It can be done using models that characterize the statistical behavior of the videotelephony sources. Videotelephony traffics consist of scenes in which people are talking with moderate motion in general and almost unchanged backgrounds [55]. Also, there are not brutal scene changes and they only occur with panning and zooming. Several models have been proposed in the literature but it was demonstrated in [61] that the GBAR(1)¹ is more specialized for modeling accurately the short-term fluctuations of single videotelephony sources and authors in [62] used the model as forecasting method in their enhanced explicit-rate mechanism.

The model relies on the observation that videotelephony traffics have gamma-marginal distributions, very high lag-1 correlation coefficient ($\rho(1) = 0.98$) and exponentially decaying autocorrelations up to lags of about 100 frames. The model is based on the fact that the sum of independent $Ga(s, \lambda)$ and $Ga(q, \lambda)$ random variables is a $Ga(s + q, \lambda)$ and the product of independent $Be(t, s - t)$ and $Ga(s, \lambda)$ random variables is a $Ga(t, \lambda)$ random variable. The forecasting rule for the model is given by:

$$X_n = A_n X_{n-1} + B_n \quad (4.4)$$

Since the distribution of X_n and X_{n-1} has to be $Ga(s, \lambda)$ as for the videotelephony traffics case, we pick A_n to be a $Be(t, s - t)$ random variable and B_n to be a $Ga(s - t, \lambda)$ random variable. We can easily see that when A_n , B_n and X_{n-1} are mutually independent, X_n is distributed as desired. The lag-1 autocorrelation function is given by $\rho(1) = \frac{t}{s}$. Using this, we can determine t since we know $\rho(1)$ and s (from the mean and variance of the data). The forecasting process is done in this way: Given X_{n-1} multiply it by A_n a sample from an independent beta-distributed random variable, and then add

¹Gamma-Beta Auto Regressive

B_n drawn from a gamma distribution. The two distributions have parameters which need to be computed only once from the mean, variance, and lag-1 correlation of the videotelephony sequence of interest.

Although its good performance, the GBAR(1) does not provide the best prediction for the frames size of the videotelephony traffics. Indeed in our previous work that we presented in [63], we used it as prediction model for videotelephony flows and we noticed that the strategy could be improved since the prediction of the frames size was not perfect and contains sometimes errors. An enhancement of the method could consist in reducing the prediction errors and improving the model.

For these reasons, we have decided in this thesis to design our own prediction model in order to improve the performance of the previous GBAR(1) prediction model and provide the best support for our SPS-P protocol. A supervised learning method has been chosen and used to build an effective traffic prediction model. Indeed, learning methods use the couples (data, label) to fit their models. In other words, they are algorithms that can learn from data and have the particularity to elaborate a model from inputs and use that to make predictions. The part below gives details on the model and show how it has been built.

a - Model description

We use the Support Vector Machines (SVMs) as machine learning algorithm to build an efficient model from a collected data in order to accurately predict the videotelephony frames size. We choose the SVMs since they are known to give good results on classification problems [64]. The main advantage of SVMs is their ability to elaborate a robust and flexible non-linear model by using parametrized kernels. In fact, SVMs are primarily classifier methods that perform classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. The process of mapping the objects from the input space to the higher dimensional feature space (called also transformation) is performed using a set of mathematical functions, known as kernels. To construct an optimal hyperplane, SVMs take advantage of an iterative training algorithm, which is used to minimize an error function. According to the form of the error function, SVMs models can be classified into four distinct forms:

- Classification SVM Type 1 (C-SVM classification);
- Classification SVM Type 2 (nu-SVM classification);
- Regression SVM Type 1 (epsilon-SVM regression);
- Regression SVM Type 2 (nu-SVM regression);

In our case, we use the C-SVM form [65] [66] [67] formally stated as follows:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + C \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & z_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, l. \end{aligned} \tag{4.5}$$

where w is the vector of coefficients, b is a constant and ξ_i represents parameters for handling nonseparable data (inputs). $\{(x_1, z_1), \dots, (x_l, z_l)\}$ is a set of training points, $x_i \in R^n$ is a feature vector, $z_i \in R^1$ is the target output, n is the number of dimensions, l is the size of the training set and $C > 0$ is the regularization parameter. We want to point out that the larger the C , the more the error is penalized. Thus, C is chosen in order to avoid over fitting. As for $\phi(\bullet)$, it is the kernel which maps \bullet into a higher dimensional space.

Vapnik-Chervonenkis theory in [67] explained that mapping inputs into a higher dimensional space (than the original dimension of the input space) often provided a greater classification power. Thus, we use Linear, Polynomial (Poly) and Radial Basis Function (RBF) kernels ϕ (which allow to map inputs into a higher dimensional space) in our experiments. They are defined as:

$$\text{Linear-kernel}(x, x') = \phi(x)^T \phi(x') = \langle x, x' \rangle \tag{4.6a}$$

$$\text{Poly-kernel}(x, x') = \phi(x)^T \phi(x') = (\gamma \langle x, x' \rangle + r)^d \tag{4.6b}$$

$$\text{RBF-kernel}(x, x') = \phi(x)^T \phi(x') = \exp(-\gamma \|x - x'\|^2) \tag{4.6c}$$

where r is a constant, d is the degree of the kernel function and $\gamma > 0$ is the coefficient for RBF and Poly kernels. It is worth noting that it is not necessary to have an explicit definition of the function ϕ , the only important thing is to have the mathematical definition of the dot product $\phi(x)^T \phi(x')$.

b - Model construction

As mentioned above, in order for the SVMs to build an efficient model to accurately predict the videotelephony frames size, they first have to learn from collected data. For this purpose we use a dataset of 8 various videotelephony trace files. These videos represent realistic videotelephony trace files (video sequences encoded using H264 constrained baseline profile with a frame rate of 25 FPS) which were downloaded from [49].

To efficiently learn models using SVM algorithm, we have to set the optimal values of hyper-parameters (which generally depends on the used dataset). Searching the optimal values for hyper-parameters is called *hyper-parameter optimization*. There are several methods of hyper-parameter optimization in the literature but in this thesis we used the *grid search* method [68]. It consists in:

- Let n be the number of hyper-parameters. For each hyper-parameter p_i ($i \in \{1, 2, \dots, n\}$), we manually define a list of (all) values that can be assigned to p_i , let l_i be this list.
- For each n -tuple (v_1, \dots, v_n) where $\forall i \in \{1, 2, \dots, n\}, v_i \in l_i$, fit a model by assigning v_i to p_i for each $i \in \{1, 2, \dots, n\}$ and assess its performance.
- We choose the hyper-parameter values that yield the highest performance.

One can notice that the model is fit and tested $\prod_{i=1}^n |l_i|$ times. Thus, the grid search suffers from the *curse of dimensionality* because the number of joint values grows exponentially with the number of hyper parameters. However it ensures that the assigned values are the most suited for our problem.

For the experiments, we used the following values for the grid search method:

- $C = [1, 10, 100, 1000]$
- kernel = [linear, poly, RBF]
- $d = [2, 3]$ (used when poly-kernel is chosen)
- label = [2, 3, 4, 5]

Finally, the optimal parameters was $C = 1$, kernel = linear and label = 5.

When using machine learning algorithms, one of the most important steps is to make sure to correctly assess the performance of the learnt model. One cannot use the training set to assess the performance of the learnt model since the performance of a model on the training set is not a good indicator for its capacity to be generalized. Therefore, the dataset must be split into 2 mutually exclusive subsets called training and test sets. In the literature, there are several methods to split datasets into training and test sets. In our case we use the *Cross-Validation* technique [69] for this task. The cross-validation method follows the steps below:

1. Randomly split the active dataset (D) into k mutually exclusive subsets (D_1, D_2, \dots, D_k) called folds of approximately equal size.

2. For each D_i ($i \in \{1, 2, \dots, k\}$):
 - Fit a model M using $D \setminus D_i$ (training set).
 - Test the model M on D_i (test set) by assessing its performance.
3. The overall performance is computed as the average performance for each fold.

k can take any value (≥ 2) but the common values are 2, 3, 5 and 10. For our experiment, since we take a value of $k = 5$, our method uses four-fifths of the dataset for the training stage and is tested on the remaining one-fifth.

Once the dataset is divided into the two sets, the SVMs algorithms are executed on the training set to learn and elaborate the model afterwards. Thereafter, the test set is used to assess the performance of the learnt model. We want to point out that the experiment is realized on the 300 first frames of the video files.

c - Model performance analysis

To show the performance of our prediction model, we compare it to the previous GBAR(1) model. The comparative study will help to highlight the ability of the learning methods to perform better than the GBAR(1) model.

The test consists in forecasting for the two prediction models the size of the 20 frames coming after the first frame of the videotelephony files located in the test set. More concretely, just the size of the first frame is given as input and the size of the next 20 frames has to be predicted. The predictions errors as well as the running time of the process are taken as key performance indicators. We consider these parameters because as explained in section 4.2.4.1, the success of the forecast process in the SPS-P protocol has two principal constraints which are time and data. Time because the eNB has to finish the forecasting and send the grants before the arrival of the second frame. Minimizing the predictions errors will help to be more accurate when allocating the RBs. We want to point out that the test has been performed on *Magi*, the computer cluster of our University. Its characteristics are presented in Table 4.1

Figure 4.3 and Figure 4.4 show for the two prediction models the prediction errors in terms of under allocation and over allocation for the 20 frames coming after the first frame of the video files located in the test set, respectively. We can observe that the SVM model outperforms the GBAR (1) model and provides best performance. Concerning the under allocation part, we can see that the SVM model provides the lowest rate. For instance the under allocation rate is reduced by 25% on the average compared to the GBAR(1) model and the confidence interval of the rate reduction at 95% is [19,98% 30,21%]. It means that for any additional frame size prediction, it is sure at 95% that

TABLE 4.1: Magi characteristics

Node type	Calculation	SMP	Login	Admin	E/S
Processor type	Intel Xeon X5670	Intel Xeon E7-4850	Intel Xeon X5670	Intel Xeon X5670	Intel Xeon X5620
Processor frequency	2.93 GHz	2 GHz	2.93 GHz	2.93 GHz	2.40 GHz
Number of cores	6	10	6	6	4
RAM	2 Go/core	512 Go	8 Go	8 Go	24 Go
Quantity	12	1	1	1	1

the rate reduction will fluctuate between 19,98% and 30,21% on the average. Clearly, the figure proves that with the SVM model, less RBs will be necessary after the pre allocation for satisfying the bandwidth requirement of the video frames belonging to the prediction window.

As for the over allocation part, we can observe that the SVM model also provides the best performance by giving the lowest rate for the prediction errors. As consequence with the SVM model, less RBs are wasted on the average compared to the GBAR(1) model. Nethertheless for the two prediction models, the prediction errors increase whenever the prediction window is extended. In other words, more the number of frames size to predict will be high, more will be the probability to make errors.

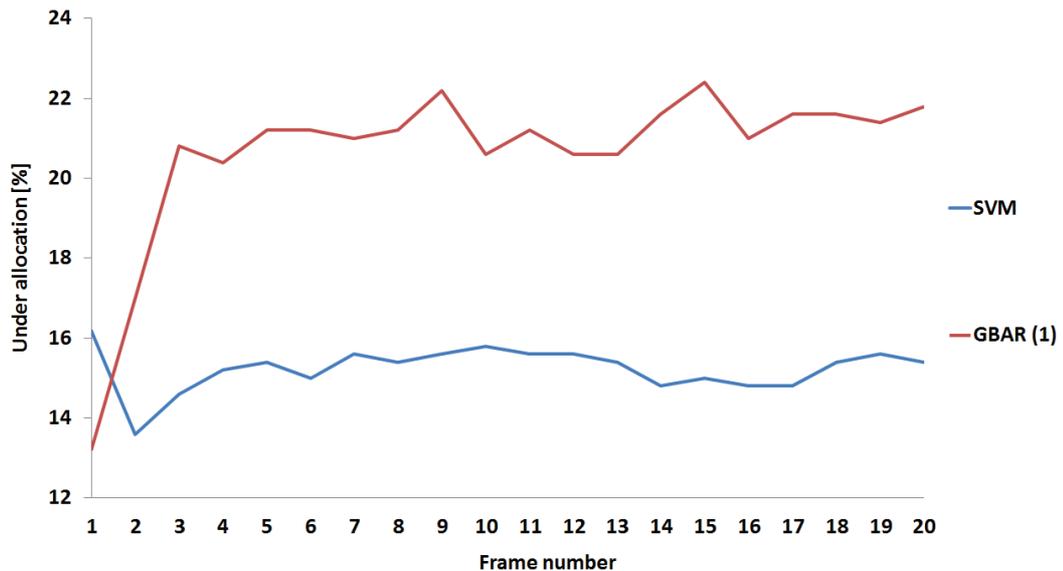


FIGURE 4.3: Prediction errors in terms of under allocation

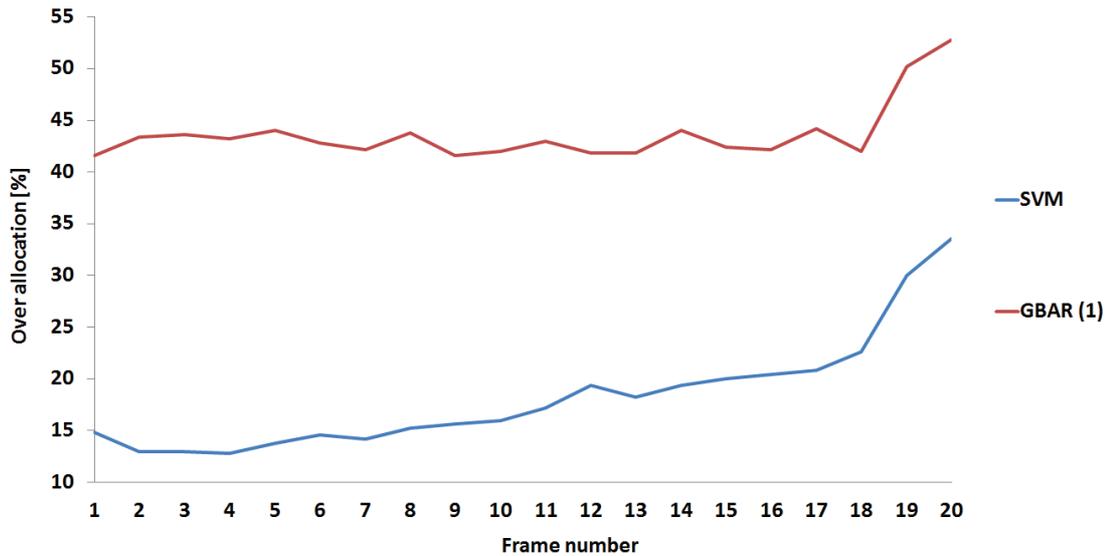


FIGURE 4.4: Prediction errors in terms of over allocation

The situation reverses in terms of the predictions running time as Figure 4.5 shows. We notice that the SVM never outperforms the GBAR(1) since the latter achieves the frames size prediction with the lowest running time. Therefore for a given period of time, the GBAR(1) model will have a larger prediction window compared to the SVM model. This is due to the fact that with the GBAR(1) model, the prediction of the frame size is computed only with the size of the previous frame. On the contrary the SVM uses the sizes of all the previous frames for the frames size prediction.

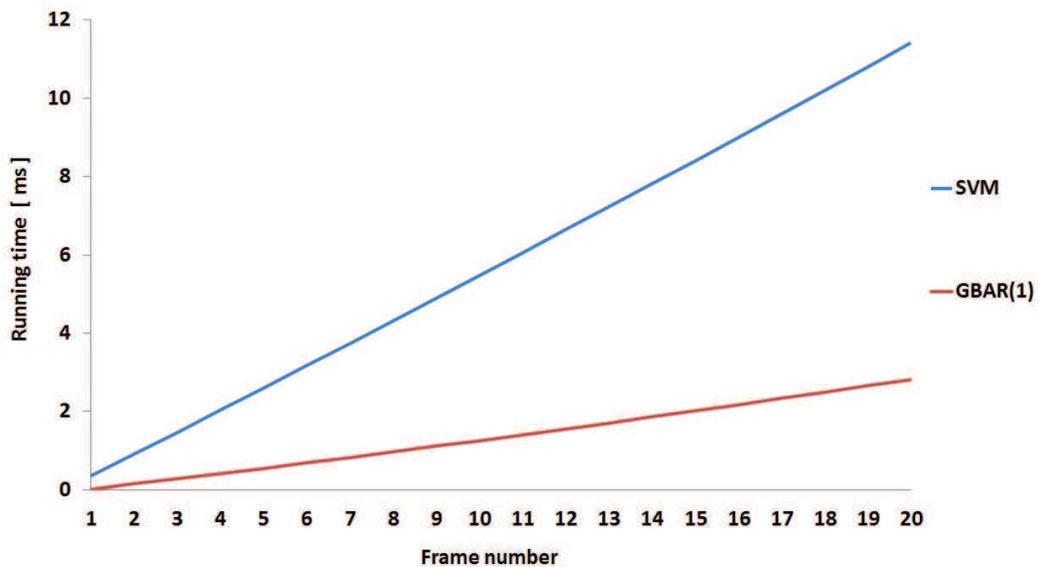


FIGURE 4.5: Predictions running time

At the end we retain that the SVM model will be used as the prediction model in our SPS-P scheme since it is better to have a short prediction window with minimal errors instead of a large prediction window with significant errors.

Knowing the amount of data to preallocate, the number of RBs or PRBs required to transmit this preallocated data is also important to determine. This is performed through the scheduling process. The part below gives a view on the scheduling algorithm.

4.2.4.3 Scheduling algorithm

The number of RBs or PRBs required to transmit the preallocated data depends both on the MCS chosen every TTI by link adaptation and on this amount of data. The eNB is obviously in charge of determining and selecting the PRBs in which each videotelephony source will transmit its predicted frames. For the TTIs where the SPS-P grants occur, the videotelephony flows have the highest priority and are scheduled first. The FME algorithm is used to assign the PRBs. The role of such algorithm is to maximize the total throughput as well as assuring fairness among flows. Once the selected PRBs cover the amount of data to preallocate to a frame of a given videotelephony user, this user is not anymore taken into account for the rest of the RBs within the SPS-P TTI. The process continues until all the videotelephony users having SPS-P grants on the considered TTI are scheduled.

In case of the PRBs required to fulfill the amount of data to preallocate to a frame of a videotelephony user is smaller than the PRBs available in the SPS-P TTI, a *non-segmentation based semi-persistent method* is adopted [70]. In this strategy, available PRBs are allocated for a part of the preallocated data and the *leftovers* are transmitted in subsequent TTIs with dynamic scheduling. For the radio resources left free in the SPS-P TTIs, they are allocated to the other users using dynamic scheduling. An overview of this method is depicted in Figure 4.6.

It is very interesting to design a new method that solves problems but it is better to study its effectiveness in realistic scenarios. This is done in the next section.

4.2.5 Performance analysis

We use LTE-Sim simulator like in Chapter 3 to evaluate the performance of our method. The simulations results aim to compare our novel scheduling protocol SPS-P using the FME allocation scheme to the Dynamic Scheduling (DS) using the same allocation scheme. These results will help to demonstrate the ability of our proposed strategy to reduce the user plane latency and provide thus, very sharp delay bounds and

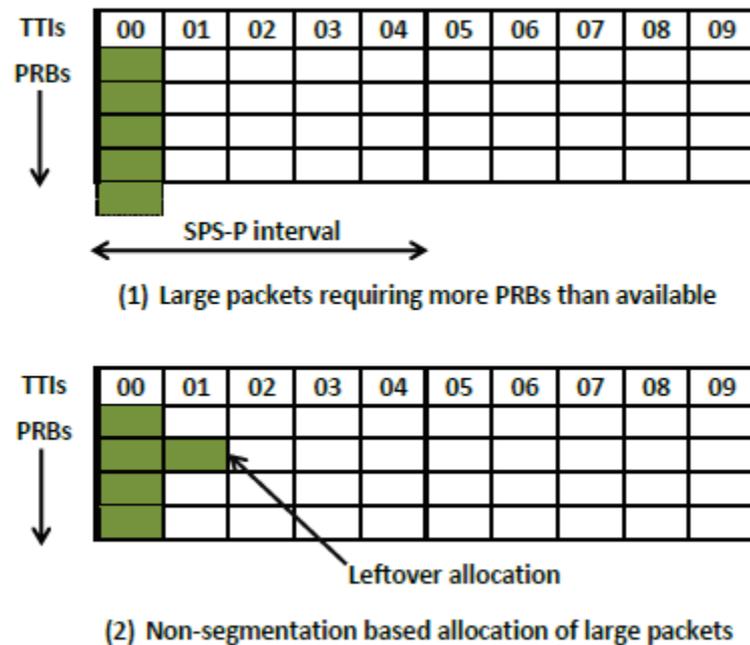


FIGURE 4.6: Non-Segmentation based allocation method

guarantee very low packet loss ratio with respect to videotelephony traffics. Therefore, we carry out simulations with different parameters which are shown in Table 4.2. The scenario consists in using a number of users of different configurations in the range [10 60] that move along in a single cell. Each user has a velocity which does not exceed 3 km/h in order to stay in the low mobility context (pedestrian). 40 % of the users handle videotelephony flows, 40 % VoIP flows and the remaining 20 % transmit best effort flows. The cumulative distribution function of the packet delays of the videotelephony flows and their packet loss ratio are taken as key performance indicators. We consider these parameters because as explained in Chapter 3, the videotelephony users have no advantages in transmitting packets after their deadline expiration. This will be a waste of resource blocks. To see the impact of our proposed scheme on the other types of users, we compute the PLR for VoIP users and the aggregate throughput for those handling best effort flows.

4.2.5.1 Traffic model

Trace-based applications are used as videotelephony flows. They send packets based on realistic videotelephony trace files (H.264) which are available on [49]. VoIP flows are G.729 flows and they are modeled with an ON/OFF Markov chain, where the ON period

TABLE 4.2: Simulation parameters

Parameters	Value
<i>General parameters</i>	
Simulation duration	50 s
Users range	from 10 up to 60
Users speed (km/h)	3
Traffic models	H.264, G.729 and Infinite buffer
Mobility model type	Random direction
Multipath model	Jakes model
<i>LTE-related parameters</i>	
Transmission bandwidth (RBs)	50
MIMO	Off
Frame structure	FDD
Cell radius	1 km
Cell number	1
Target System	Uplink
Schedulers type	SPS-P and DS
Max delay	0.1 s
Target System	Uplink
Schedulers type	SPS-P and DS
Max delay	0.1 s
Prediction window N_f	10 frames
Number N_i	3
Threshold S_t	80 %

is exponentially distributed with a mean value of 3 s and the OFF period has a truncated exponential probability density function with an upper limit of 6.9 s and an average value of 3 s [50]. During the ON period, the source sends 20 bytes sized packets every 20 ms (the source data rate is 8.4 kbps), while during the OFF period the rate is zero because the presence of a VAD is assumed. Best effort flows are created by an infinite buffer application which is modeled like an ideal greedy source that always has packets to send. In the part below, we present the simulation results and give our analyses.

4.2.5.2 Results and analyses

As shown in Figure 4.7 and Figure 4.8, our method outperforms the dynamic scheduling strategy by always reaching the lowest value for the PLR and by giving the best performance for the CDF of the packet delays of videotelephony flows.

Regarding the PLR, we can observe that our proposed protocol permits to reduce it by 71,93 % on the average. Concerning the CDF of packets delays, the difference between the two methods is more obvious in the case of high traffic load. For instance when 60 users are in the cell, 90 % of the videotelephony packets scheduled with our algorithm has a delay lower than 50 ms, whereas just 50 % of the videotelephony packets scheduled with the dynamic scheduling has a delay lower than 50 ms. It is just because with our strategy, videotelephony users do not have to wait before transmitting packets. Indeed, once the video frames belonging to the prediction window are available in the buffer, they are directly transmitted without any procedure because resources have been already preallocated for them. For the dynamic scheduling, it is different because resources are not guaranteed. Therefore, the videotelephony users can have significative user plane latency and thus, a great number of expired packets in their queues.

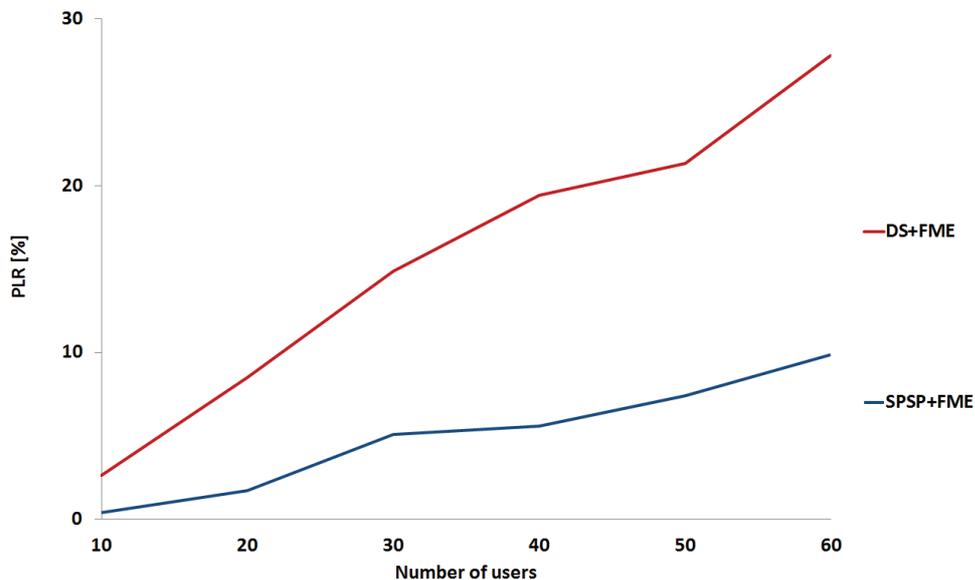


FIGURE 4.7: PLR of videotelephony users

To study the impact of our scheme upon the other types of UEs, Figure 4.9 presents the PLR for VoIP users and Figure 4.10 depicts the aggregate throughput for those which hold best effort flows. In these figures, we can observe that our scheme does not have the best performance. It can be explained by the fact that the priority given to videotelephony users has the cost of reducing the amount of resources to allocate to

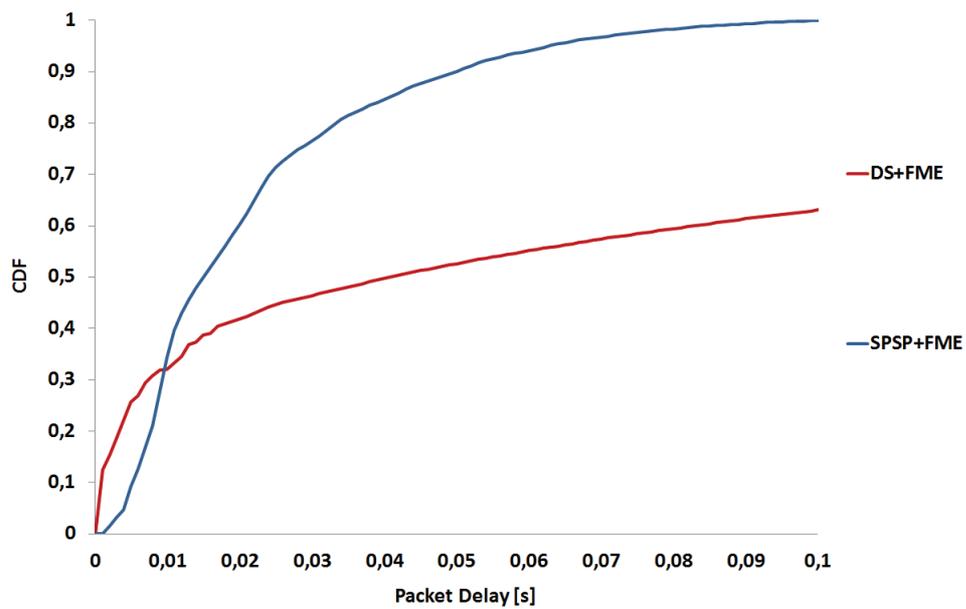


FIGURE 4.8: CDF of video packet delays with 60 users

the other UEs. Nevertheless with dynamic scheduling it is different because no priority is given to videotelephony users. Therefore, the other types of users can have more resources. This is why with the dynamic scheduling, the obtained PLR for the VoIP users is the lowest one and the best effort users are given more resources.

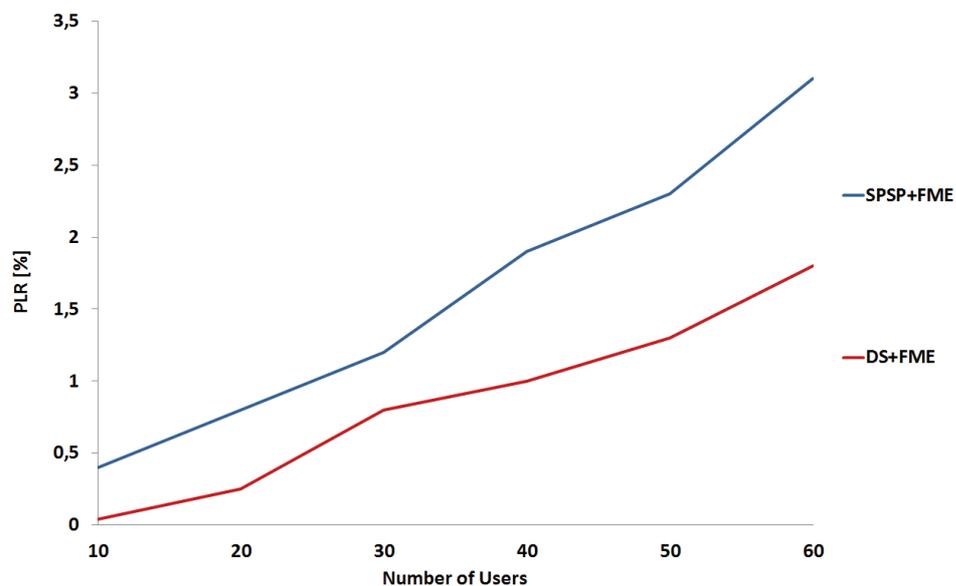


FIGURE 4.9: PLR of VoIP users

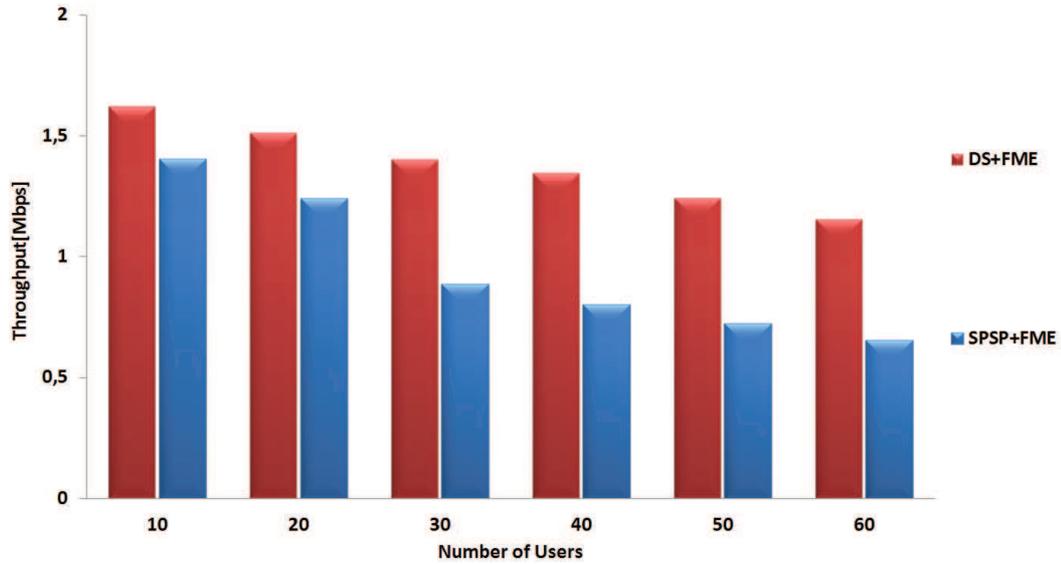


FIGURE 4.10: Aggregate throughput of Best Effort users

4.3 Conclusion

In Chapter 3, we focused on the downlink direction and proposed efficient strategies able to improve the scheduling process in order to provide the best QoS support to the RT flows. However as detailed at the end of this Chapter, the designing of efficient resources allocation must be also extended to uplink direction. Therefore throughout this chapter, we addressed the problem of uplink resources allocation in a context where the high mobility of the users is not taken into account. Designing efficient resources allocation strategies in the uplink direction is more difficult since the proposed method should not only deal with the constraints related to the scheduling process, but should also be compatible with the constraints required by this direction. Among these constraints, the allocated RBs must be contiguous and the scheduler must be continuously informed by the UEs about their amount of data to be transmitted (the eNB is not anymore aware of UEs data awaiting in the buffer like in the downlink). These constraints lead to modify the scheduling process and also to increase the user plane latency. Indeed, we have seen that the handshake procedure consisting of a scheduling request message from the UE and a scheduling grant from the eNB required twice a communication over the air interface and caused notable delay. We have demonstrated that this delay could exceed 12 ms for the case of videotelephony traffics. As consequence, the resources allocation for these types of traffics could lead to weak QoS support.

Therefore, we proposed a new scheme to improve resources allocation for these traffics and reduce the delay caused by dynamic scheduling. It is based on the design and the use of a new scheduling protocol for handling differently videotelephony traffics in the

uplink of LTE networks. The key idea consists in scheduling such traffics using a semi-persistent method with a provisioning process (SPS-P). In fact, VBR videotelephony traffics in the network are characterized by variable sized packets with constant inter-arrival time. With the semi-persistent mode, the UEs do not request anymore RBs for their video frames. As for the provisioning mode, it helps to define the amount of resources to reserve for each frame by using an accurate traffic prediction model. A supervised learning method has been used to elaborate an effective traffic prediction model. Indeed, learning methods are algorithms that can learn from data and have the particularity to build a model from inputs and use that to make predictions. In our case, we used the SVM to learn from a dataset of various videotelephony trace files. We have seen that the obtained traffic prediction model was able to provide the best performance compared to the previous GBAR(1) model. For instance, the SVM performed the frames size forecasting with the lowest rate of prediction errors.

The proposed SPS-P strategy have been evaluated in real LTE environments with LTE-Sim and simulations results proved its effectiveness by showing that it improved videotelephony traffics performance and provided the best QoS support. Indeed, the simulations results demonstrated that our scheme strongly reduced the packet loss ratio of the videotelephony flows and provided the lowest value for their packet delays.

Throughout Chapter 3 and Chapter 4, we designed efficient scheduling schemes which provided the best performance compared to several ones in the literature. They helped to improve both for the downlink and the uplink, the QoS of RT flows in a context where the high mobility was not taken into account. Since we have acquired an expert knowledge of the scheduling process in such context, we now focus on designing suitable methods for the high mobility scenarios. In other words, in next chapter we will address the problem of resources allocation in a context where the high mobility of the UEs is taken into account. This is the case of vehicles equipped with their own communication interfaces. For this type of nodes, the network provides internet access and enables also the use of applications which aims for instance at increasing the passengers safety, at improving also the driving experience or at providing entertainments. Nethertheless, their high mobility aspect due to their high velocity can affect the resources scheduling process and lead to performance degradations [9]. Therefore, the network could not be able to provide strong QoS support to the applications of this type of users. In order for the network to provide the best QoS support, it is important and necessary to design effective scheduling schemes adapted to the high velocity scenarios.

In the next chapter, we present our novel methods designed for improving resources allocation in these scenarios.

Chapter 5

Resources allocation in high mobility scenarios

This chapter presents our major contributions to face the problem of resources allocation in high mobility scenarios of LTE networks. We have firstly proposed a technique which maintains the required level of QoS for supporting video users at high velocities. This technique consists in identifying depending on the UEs velocity, the minimum CQI reports rate in order to maintain the required QoS for supporting the users video flows. Secondly, we designed an opportunistic method which improves the performance of high speed video users in vehicular scenarios. Indeed, it appeared from our study on the impacts of high mobility scenarios that performances of high velocity users were the most affected. In fact the higher was the speed of the UEs, the worse were their performance. Therefore with the proposed strategy, users with the highest velocity are given more priority and resources, thereby helping them to increase their performance. Simulations results have confirmed its effectiveness and showed that the proposed scheme helped to improve the performance of the video users having the highest velocity by reducing the PLR and delay of their flows.

5.1 Context

Throughout Chapter 3 and Chapter 4 of this thesis, we addressed the problem of resources allocation in the downlink and the uplink respectively. In these parts, we designed and implemented efficient scheduling schemes for improving QoS of RT applications. These strategies have proved their effectiveness and provided the best performance compared to those in the literature. Nethertheless the high mobility aspect

of the UEs was not taken into account when designing these algorithms. In these chapter, we consider this mobility aspect to propose suitable schemes.

Indeed with further advances in high speed conditions, the support of high data rate and voice communications for high velocity users has become urgent demands [71]. Therefore, effective mobile network must support users having high velocities and bring them strong QoS. LTE has been considered as one of the technologies to support the high velocity users. This is the case of vehicles equipped with their own communication interfaces. For this type of nodes, the network must provide for instance strong QoS support to their applications which aims at increasing the passengers safety, at improving also the driving experience or at providing entertainments.

Nethertheless, their high mobility aspect due to their high velocity can affect the resources scheduling process and lead to performance degradations [9]. Indeed in high speed environments, the channel measurement for channel estimation may not be able to adapt fast enough to channel variation [72]. This results in scheduling errors and lead to a performance degradation. As consequence the network could not be able to provide the best QoS support to this type of users.

For this purpose, we propose two efficient strategies adapted to these scenarios. The vehicular environment and the video flows are considered. We focus on vehicular scenarios since they represent a significant challenge for networks operators, given the combination of their large data volumes, elevate speed and unique movement patterns. As for video flows, they are the most used applications as explained in Chapter 4. The first proposed scheme is a technique which maintains the required level of QoS for supporting video users at high velocities. The second one is an opportunistic scheduling method which improves the performance of high speed video users by considerably reducing their performance degradation.

Through the study on the impact of the effects involved with the vehicular scenarios, we highlight their consequences on the users performance. Based on the results of this study, we suggest our tailored solutions to meet the needs.

5.2 Vehicular scenarios overview

Recently, vehicular scenarios in mobile networks have attracted a considerable attention from the research community since the aim of connecting a car to the network is to provide not only safety and efficiency in the vehicle but also leisure and infotainment with the driver and the passengers [73] [74].

In such scenarios, the vehicles are active mobile users which are equipped with interfaces enabling connection and communications with the access network. An example for the LTE case is depicted in Figure 5.1.

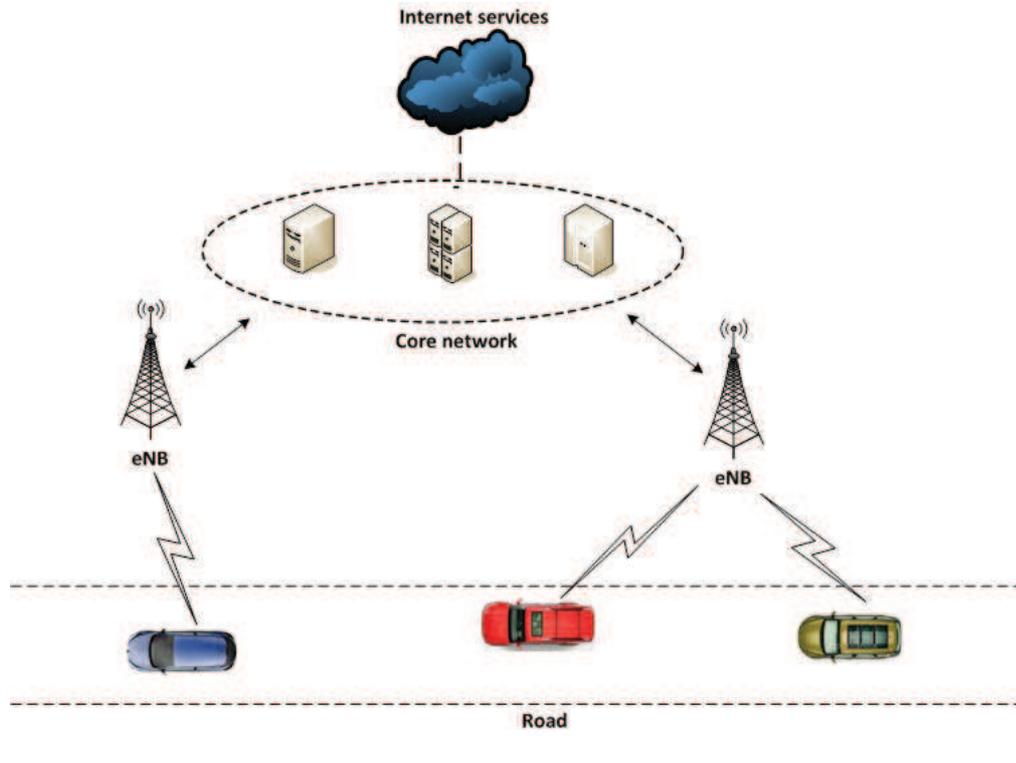


FIGURE 5.1: Generic view of vehicular scenarios in LTE network

The connection and communications with the network have two main purposes. On the one hand, the network is used to provide seamless internet connectivity with the on-board passengers. On the other hand, it is used to upload the so-called Floating Car Data (FCD). Actually the FCD consists in data which is generated periodically by the vehicles in an autonomous fashion and transferred to data centers for processing via the mobile network infrastructure [75]. Figure 5.2 illustrates an example. Practical usages of FCD include RT road traffic monitoring, fleet management and distant support for safety, diagnostic and anti-theft services [76] [77].

Clearly, the aim of the connected car concept is to make the time spent in vehicle more enjoyable to both drivers and passengers and improve the road traffic safety and efficiency. However managing vehicular scenarios is very challenging for the mobile network since it has to deal with a specific characteristic which is the high velocity of the nodes. This constraint has significant impacts on network performance and needs to be taken into account when designing any scheme for bringing strong QoS support in these scenarios.

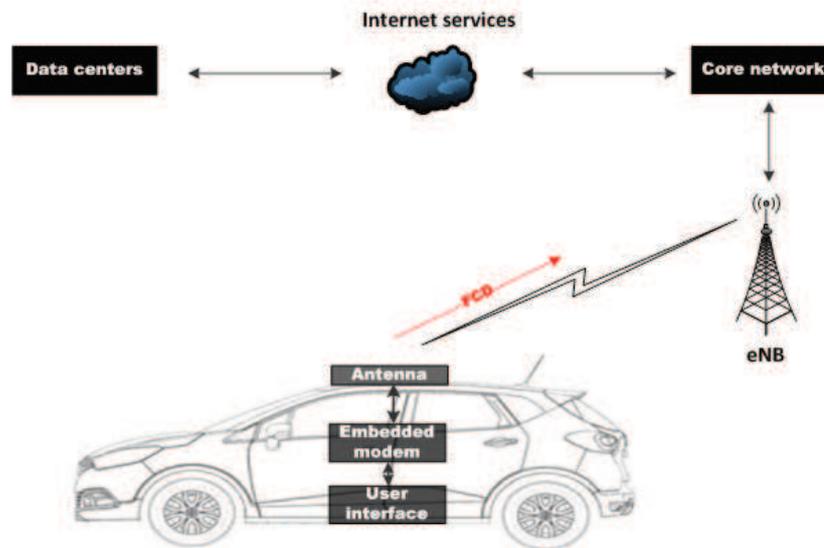


FIGURE 5.2: Example of FCD uploading via LTE network

In next part we highlight the consequences of high velocity scenarios on users performance by firstly analyzing the effects involved with these scenarios and after quantifying their impacts.

5.3 Effects involved with the high velocity scenarios

Basically, there are two main effects that can be observed at high velocity scenarios: On the one hand, the transition from one cell to another one becomes more frequent. This leads to higher handover rates. On the other hand, the channel conditions change faster and become generally more unstable [78].

Regarding the high handover rate, it does not have a significant impact on the LTE system performance. Indeed the seamless and lossless handover mechanisms implemented in LTE network helps to avoid significant system performance degradations. Several papers have shown that LTE can achieve good handover performance in terms of user throughput, handover delay and handover failure rate, even with higher handover rates [79] [80] [81]. One of the major reasons is the low latency in the X2 interface (links between eNBs) and the fast processing within the eNBs. The low latency in the RAN suppresses the risk of the node losing its connection with the serving cell, while still waiting for the handover command.

As for the faster changing channel conditions, it is different. This effect involved with the higher nodes speeds has a negative impact on the LTE system performance. Indeed as seen in Chapter 2, most of schedulers in LTE consider the channel state in

terms of CQI reports from UEs to efficiently utilize the limited radio resources. In other words their scheduling metrics take into account information about the channel quality. However with the faster changing channel conditions due to the high velocity, the channel measurements for channel estimation are not able to adapt fast enough to channel variation [72] [82]. Actually the current channel conditions at the actual scheduling instant significantly deviates from the reported channel quality. Then CQI reports become not only outdated sooner but also more and more unreliable.

The erroneous feedbacks provided by the nodes affect the metric calculation and result in errors in scheduling decisions. As consequences, users performances are subject to degradation [83]. In the following part, we go deeper in this issue and show its concrete impact on the users performance.

5.4 Impact of the effects involved with the high velocity scenarios

In order to quantify the impact of the effects involved with the high velocity scenarios, we performed several simulations with LTE-Sim. The principles of the used simulator are the same as presented in Chapter 3 and Chapter 4. The most important simulation parameters are presented in Table 5.1 below.

The simulations were performed for two types of scenarios which descriptions are shown in Table 5.2.

Scenario 1 consists in using a number of users in the range [12 72] that move along in a multi-cell urban environment. The UEs do not have the same configuration during the simulations in terms of velocity. Indeed, the velocities (3, 30 and 60) are uniformly distributed among the users in order that the UEs have different velocities during the simulation. As for scenario 2, the simulation is carried out in a multi-cell rural environment. For this case, a fourth velocity (90 km/h) is added within the simulation and the velocities are uniformly distributed among the UEs as for scenario 1.

Within all the scenarios, each UE receives a video flow. The video flows used have the same characteristics than that presented in Chapter 3 and Chapter 4. In these simulations we target the downlink direction but most of the results and lessons learned hold for the uplink as well. We choose our Enhanced Frame Level Scheduler (E-FLS) algorithm since it has provided the best performance in the downlink when scheduling multimedia flows (See part 3.3).

TABLE 5.1: Main simulation parameters

Parameters	Value
<i>General parameters</i>	
Simulation duration	50 s
Users Range	from 12 up to 72
Users Speed (km/h)	3, 30, 60 and 90
Traffic models	H.264 (Video)
Channel model type	Vehicular A
<i>LTE-related parameters</i>	
Transmission bandwidth (RBs)	50
MIMO	Off
Frame structure	FDD
Cell radius	1 km (Urban) and 5 km (Rural)
Target System	Downlink
Scheduler type	Enhanced frame level scheduler (E-FLS)
CQI reporting interval	Periodic (1 ms)

For each UEs configuration, we measure the CDF of the packet delays of the video flows and their PLR. We consider these parameters because RT applications like video flows have no advantages in receiving expired packets and transmitting them after the deadline expiration. This will be a waste of RBs. We evaluate also the throughput (average) achieved by these flows.

TABLE 5.2: Simulation cases

Case	Scenario type	Speeds used in the scenario	Description
Scenario 1	Multi-cell urban	3 kmph, 30 kmph and 60 kmph at the same time	The different velocities are uniformly distributed among the users
Scenario 2	Multi-cell rural	3 kmph, 30 kmph, 60 kmph and 90 kmph at the same time	The different velocities are uniformly distributed among the users

Figure 5.3 and Figure 5.4 show the PLR of video flows under several traffic loads for the different simulation cases. At a first glance, it can be seen that the PLR increases

with the traffic load. In addition, we notice that the higher is the speed of the UEs, the higher is the PLR of their video flows. At this level, there is a significant performance degradation for the users having the highest speed (60 km/h for urban and 90 km/h for rural) compared to the those having a velocity of 3 km/h. For instance for scenario 1 the PLR for is increased by 128,94 % on the average. This can be explained by the fact that most of schedulers in LTE like E-FLS algorithm that we used within these simulations, consider the channel state in terms of CQI reports from UEs in their scheduling metrics. For the high velocity users, the part of the scheduling metric containing the channel quality is affected by the erroneous feedbacks provided and thus the metric calculation impacts their priority.

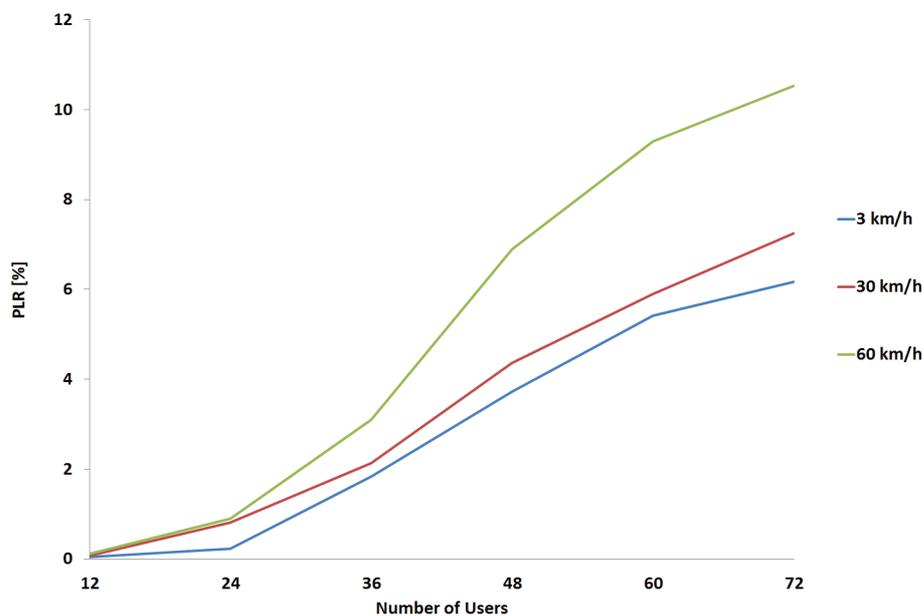


FIGURE 5.3: PLR of video flows under several traffic loads for scenario 1

The performances degradation of the high velocity users is also demonstrated with the CDF of the packet delays of video flows and the average throughput obtained by these users.

Regarding the CDF of the packet delays, we can see without any surprise in Figure 5.5 that low velocity users achieve the best performance whereas the high velocity users achieve the worst. Indeed, packets of the UEs of 3 km/h have the lowest delay compared to those of 60 km/h which possess the highest. As for the average throughput per user, we can observe that this throughput decreases with the increase of the speed as depicted in Figure 5.6. The difference is more significant at high traffic load. These performance degradation can be explained by two main reasons. On the one hand, with high velocity more erroneous feedbacks are provided and this impact the priority of the high velocity

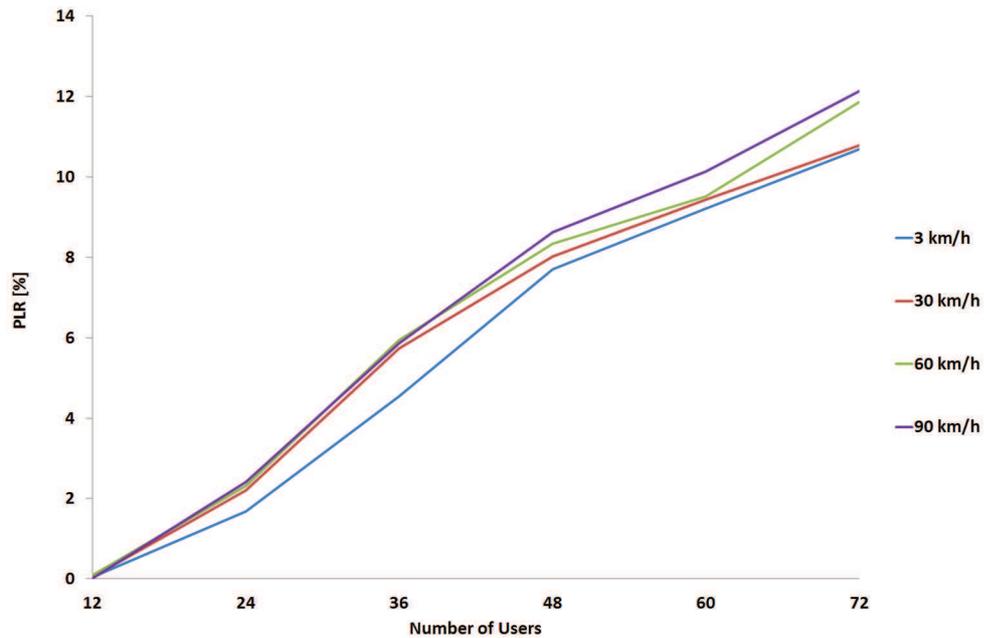


FIGURE 5.4: PLR of video flows under several traffic loads for scenario 2

users. On the other hand increased velocity affects on decreasing the selection of MCS and usage of lower ones which lead to lower throughput.

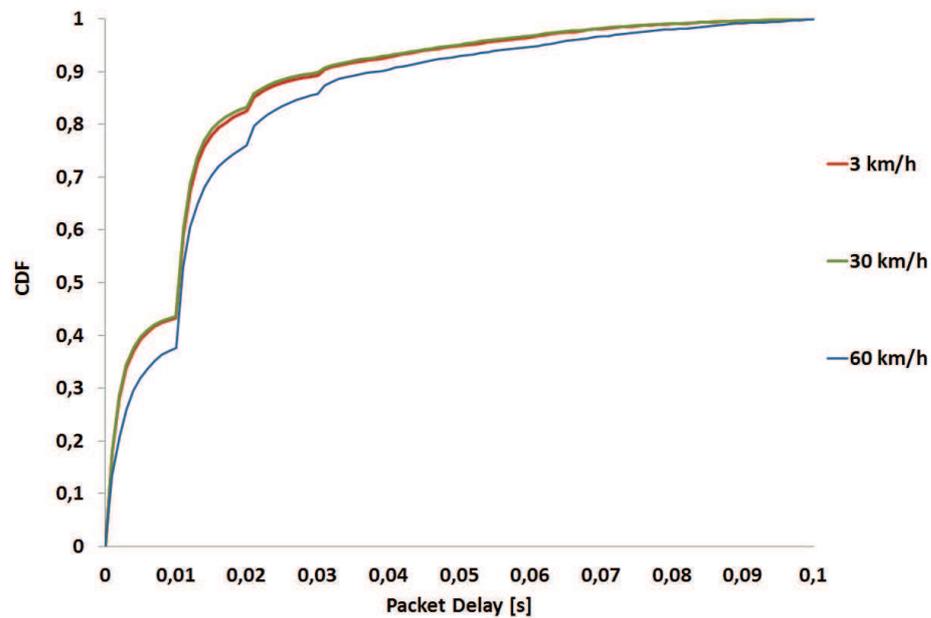


FIGURE 5.5: CDF of video packet delays with 60 users for scenario 1

From these simulations, we conclude that the high velocity scenarios have a negative impact on the users performance. Indeed we noticed a performance degradation with

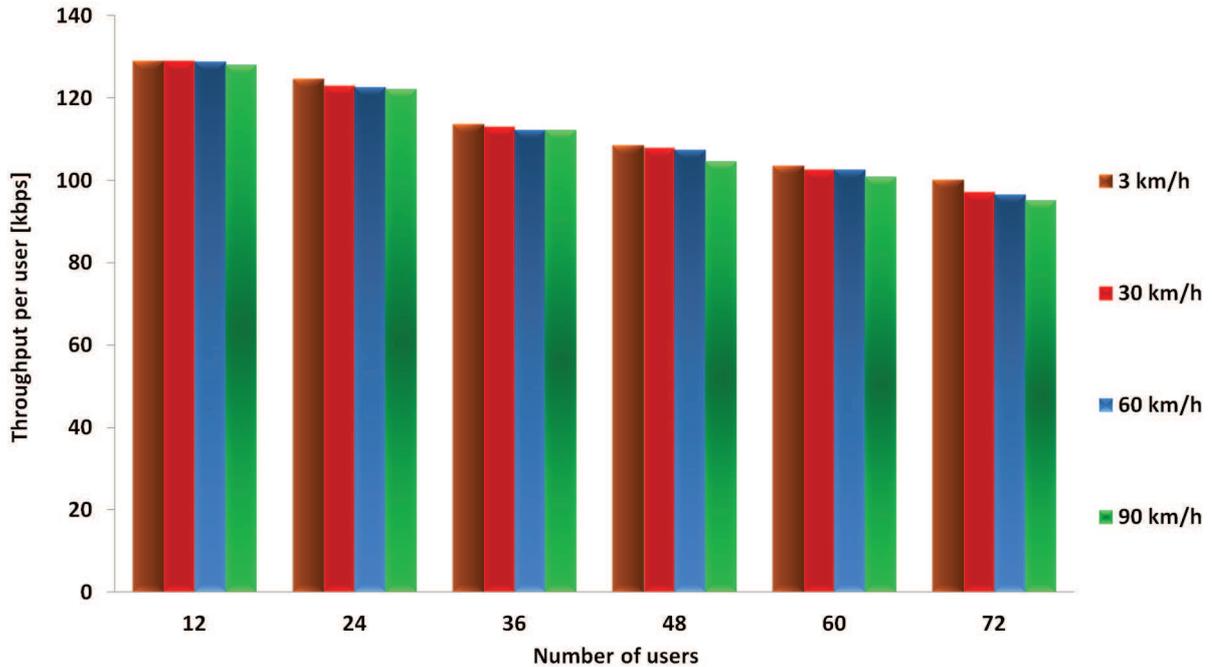


FIGURE 5.6: Average throughput of video flows for scenario 2

the increase of the speed. Within all the simulations, the high velocity UEs were the most affected and obtained significant performance degradation. The reason is that channel conditions change faster and become more unstable at high speed. Thus the channel measurements for channel estimation are not able to adapt fast enough to channel variation. Then the part of the scheduling metric containing the channel quality is affected by the erroneous feedbacks provided. As consequences errors occur in scheduling decisions and impact the users scheduling priority.

Clearly the high velocity scenarios will lead to the UEs performance degradation and cause deterioration of the QoS of their applications. However LTE has been designed to bring strong QoS support even to high mobility users. Therefore in order to reach this objective, it is important and necessary to design appropriate schemes adapted to these scenarios. The following section presents our contributions.

5.5 Proposed methods

As explained in Chapter 4, video flows have become the most used applications and it is obvious that with the connected car concept and the LTE network capabilities, their usage is going to be even higher and the network has to provide strong QoS support for those services. However due to the effects involved with the high velocity scenarios,

it will be very challenging to provide in these scenarios the required level of QoS and maintain the planned system performance [84]. Indeed, we observed from our study on the impacts of such scenarios that they lead to the users performance degradation. The higher was the speed of the users, the worse were their performance. This could be harmful for video applications within those scenarios.

Therefore it is necessary and important to design effective schemes adapted to the high mobility scenarios of LTE networks. These schemes could permit to improve the QoS support for the high speed video users. For this reason, we propose two efficient strategies. The first one is a technique which maintains the required level of QoS for supporting video users at high velocities. The second one is an opportunistic scheduling method which improves the performance of high speed video users by reducing their performance degradation.

5.5.1 CQI rate compromise

In LTE, most of schedulers exploit the channel state in terms of CQI reports from UEs to perform the scheduling process. However with the faster changing channel conditions due to the high speed environment, the channel measurements for channel estimation are not able to adapt fast enough to channel variation and the current channel conditions at the actual scheduling instant significantly deviates from the reported channel quality. As consequences the CQI reports become not only outdated sooner but also more and more unreliable. These erroneous and outdated feedbacks provided affect the scheduling metric and result in errors in scheduling decisions and hence degradation of the users performances. Based on these considerations, one possible solution to avoid the performance degradation would be to reduce errors in the CQI received by the eNB. One could want to increase the CQI reports rate from the UEs in order to reduce the probability of errors. However, this leads to increase the uplink signaling overhead. We notice that the amount of resources dedicated to the signaling is limited, thus increasing the CQI rate can decrease the degrees of freedom for the scheduler.

So our proposed solution consists in doing a trade-off between uplink signaling overhead and improving the scheduling performance. In other words the key idea consists in identifying based on the UEs velocity, the minimum CQI reports rate to maintain the required QoS for supporting the users applications. For this purpose, we perform several simulations with LTE-Sim using the same simulations parameters as in section 5.4. The simulation scenario consists in analyzing the PLR of video flows by varying the CQI reports interval (CRI) from 72 UEs having the same velocity configuration as those in scenario 2.

A similar approach was proposed in [84] and it was so relevant to us. However all the simulations were carried out with the M-LWDF algorithm and only with users at two velocities. In our case we carry the analysis further and consider an environment with several velocities and with the E-FLS algorithm which shows better performance for video flows compared to M-LWDF algorithm.

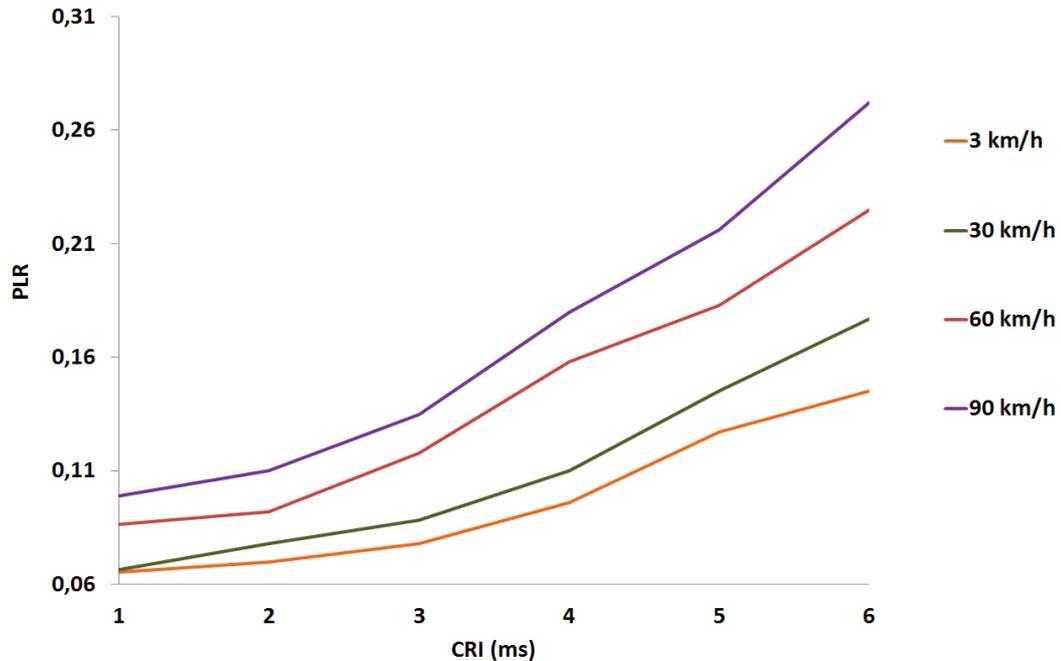


FIGURE 5.7: Packet loss ratio of video flows under various CQI reports interval

The results of the experiment are presented in Figure 5.7. It can be seen that the PLR increases whenever the CRI increases. We notice also that the performance degradation is more pronounced at the high velocity users side. The reason is just because as the CRI increases, the eNB is restricted of more and more up-to-date information and hence increasing the PLR of video flows.

To apply our proposed solution in this experiment, we firstly target the level of PLR to maintain, and then derive the appropriate CRIs for the different types of users. So we can see in the Figure that to maintain a PLR of 0,1 for the video flows, 1 CQI per TTI is necessary for the UEs going at 90 km/h. However for those going at 3 km/h, 1 CQI every 4 TTIs is approximately sufficient. Using the same method, we can find also the appropriate CRI for the other performance metrics of the video flows such as the users throughput and delay.

Clearly by this process we can identify the CRI needed by the different types of users in order to maintain the required QoS for the video flows. Nevertheless this method

becomes ineffective if we want to improve the performance of the high velocity UEs instead. The following part presents the proposed strategy designed to meet this need.

5.5.2 Opportunistic scheduling scheme

5.5.2.1 Context and key design aspects

We concluded from the simulation results obtained in section 5.4 that the high velocity scenarios had a negative impact on the users performance. We observed that the UEs with the highest velocity were the most affected and underwent significant performance degradation within all simulations compared to the other UEs. In order to avoid weak QoS support to RT flows users, we proposed a technique which helps to maintain the required level of QoS of their application. However this method becomes inefficient if one want to bring strong QoS support to their applications, ie improve the performance of these high velocity UEs. For this purpose, we introduce an opportunistic scheme which allows to increase performance of the users with the highest velocity. The idea of this proposition relies on the fact that in high velocity scenarios like vehicular scenarios or VANETS (Vehicular Adhoc NETWORKS), it is advised to give more privileges to the UEs with the highest velocity since they spend less time in the network compared to the other [85].

To reach this aim, the strategy to design must give more priority and resources to the UEs with the highest velocity. This can be achieved by defining a new scheduling scheme adapted to this situation. In the literature most of the proposed schedulers for LTE networks do not consider the high mobility user scenario. The only one that we found in [83] proposed an extended version of the PF scheme which takes into account the users velocity and provides better performance in terms of fairness and UEs throughput compared to the original version of the PF scheme. Although its originality, this strategy cannot be efficient for multimedia flows which require strict constraints on packet delay and packet loss. Indeed, we have shown previously (see Chapter 2 and Chapter 3) that PF strategy does not provide good performance for RT flows. In fact it is not a QoS-aware scheme. Thus to the best of our knowledge, an efficient scheduling scheme able to improve performance of high velocity video users in LTE networks still has to come.

For this purpose, we design an opportunistic scheduling strategy which improves the performance of high speed video users by considerably reducing their performance degradation. The scheme is the modified version of our E-FLS algorithm adapted to the high velocity scenarios. With this scheme, the scheduler takes into account the users velocity to allocatate the RBs. The metric of this policy is expressed by:

$$M_{HighVelocityScheduler} = M_{EFLS,i} \cdot \log_2[V_i] \quad (5.1)$$

with:

$$M_{EFLS,i} = \text{Exp}\left(\frac{a_i \cdot D_{HOL,i}}{1 + \sqrt{\frac{1}{N} \sum_t D_{HOL,i}}}\right) \cdot \frac{R_{i,k}}{\bar{R}_i} \quad (5.2)$$

and

$$a_i = -\frac{\log \delta_i}{\tau_i} \quad (5.3)$$

where V_i is the velocity of the UE holding the i -th flow. We want to point out that in LTE networks, a user velocity can be estimated using several techniques such as A-GNSS (Assisted Global Navigation Satellite System), enhanced cell-ID positioning and OTDOA (Observed Time Difference Of Arrival) which has been introduced specially for the release 9 (See Appendix B for description).

τ_i is the largest delay that packets of the i -th flow can tolerate and δ_i is the largest probability with which the delay requirement can be violated. $D_{HOL,i}$ represents the head of line packet delay of the i -th flow and $R_{i,k}$ and \bar{R}_i are the expected data rate for the i -th flow on the k -th RB and the past average rate achieved by the i -th flow respectively. N is the total number of active RT flows.

The presence of V_i in the scheduling metric helps to give the users with the highest velocity the possibility to acquire more resources in order to increase their performance. As for the logarithm function, it allows to mitigate the differences between the users velocities.

It is very interesting to design a new method that solves problems but it is better to study its effectiveness in realistic scenarios. This is done in the following part.

5.5.2.2 Performance analysis

We use LTE-Sim simulator as in the previous simulations to evaluate the performance of our proposed strategy. The simulations results aim to compare our proposed scheme to the Enhanced Frame Level Scheduler (E-FLS). These results will help to demonstrate the ability of our strategy to improve the performance of high speed video users. Thus, we perform several simulations with different parameters which are shown

in Table 5.3. The scenario consists in using a number of users in the range [20 120] that move along in a rural environment. The velocities (3, 30, 60 and 90) are uniformly distributed among the users so that the UEs have different velocities during the simulation. Each UE receives a video flow which has the same characteristics than that presented in Chapter 3 and Chapter 4. We measure for the highest velocity users (90 kmph), the CDF of the packet delays of their video flows, their average delay and their PLR.

TABLE 5.3: Simulation parameters

Parameters	Value
<i>General parameters</i>	
Simulation duration	50 s
Users range	from 20 up to 120
Users speed (km/h)	3, 30, 60 and 90
Traffic models	H.264 (Video)
Channel model type	Vehicular A
<i>LTE-related parameters</i>	
Transmission bandwidth (RBs)	50
MIMO	Off
Frame structure	FDD
Cell number	1
Cell radius	5 km
Scenario type	Rural
Target System	Downlink
CQI reporting interval	Periodic (1 ms)

As depicted in Figure 5.8 and Figure 5.9, our strategy outperforms the E-FLS algorithm by always reaching the lowest value for the PLR and by providing the best performance for the CDF of the packet delays of videos flows. For instance with our method the PLR is reduced by 15,58% on the average and the confidence interval of the PLR reduction at 95% is [10,47% 20,69%], meaning that for any additional simulation which will be done outside the traffic load interval, it is sure at 95% that the PLR reduction will fluctuate between 10,47% and 20,69% on the average.

To confirm the performance of our scheme concerning the delay, we also depict for the 90 kmph users in Figure 5.10, the average delay of their video flows scheduled with the E-FLS algorithm and with our proposed scheme. We can observe that our scheme

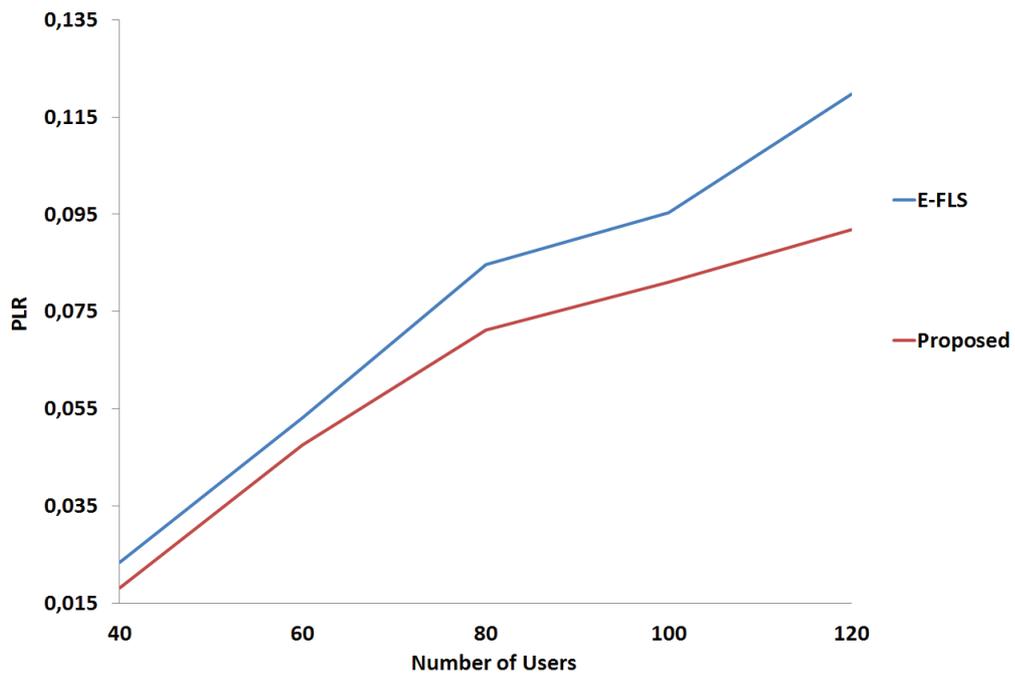


FIGURE 5.8: PLR of video flows of 90 kmph users under various traffic loads

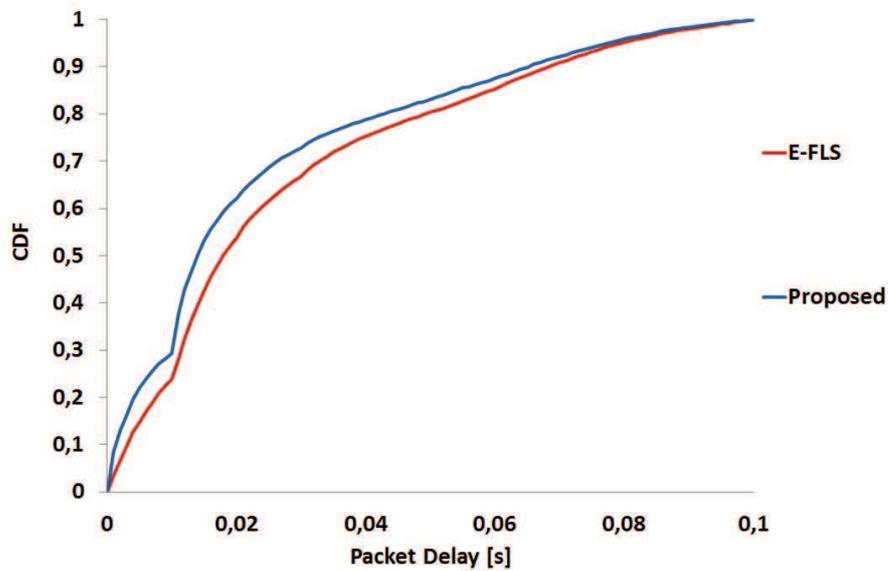


FIGURE 5.9: CDF of video packet delays of 90 kmph users

which is a modified version of the E-FLS, outperforms it by reducing efficiently the average delay of these flows under several traffic loads. As for the PLR case, the delay is also reduced by approximately 15% on the average. The reason is that with our method, the users with the highest velocity are given more priority and a higher number of RBs compared to the E-FLS case where they are not prioritized. This has the consequence

to improve their performance and thus, lead to reduce the PLR and the delay of their video flows.

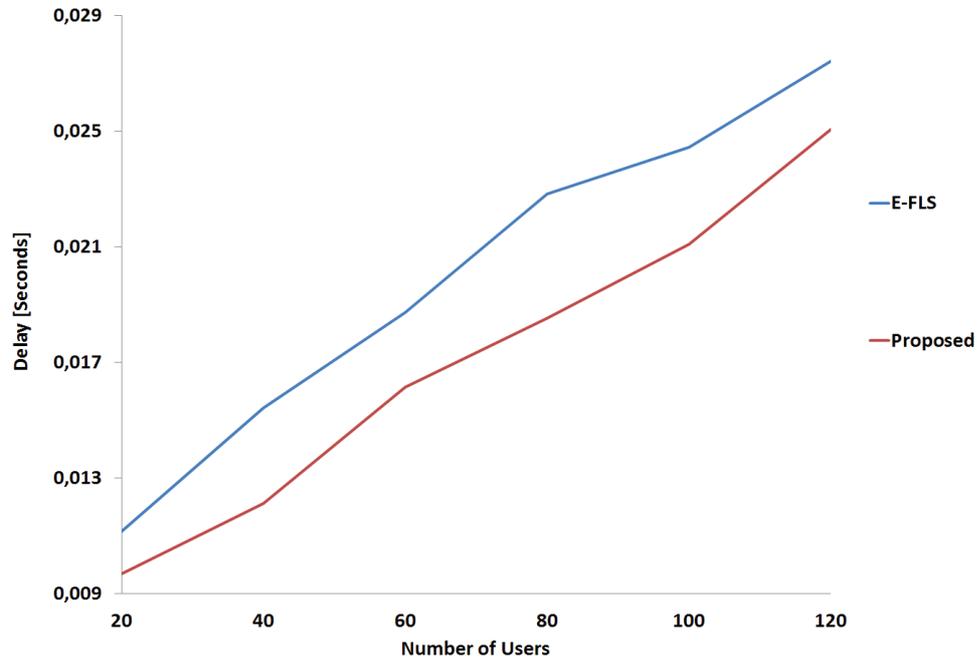


FIGURE 5.10: Average delay of video flows of 90 km/h users under several traffic loads

5.6 Conclusion

Throughout this chapter, we addressed the problem of resources allocation in a high mobility context. We focused on the case of vehicular scenarios since they represent a significant challenge for networks operators, given the combination of their large data volumes, elevate speed and unique movement patterns. In these scenarios, the nodes are vehicles equipped with their own communication interfaces. For this type of nodes, the network must provide internet access and enable also the use of applications which aims for instance at increasing the passengers safety, at improving also the driving experience or at providing entertainments. It consists in designing efficient scheduling schemes which takes into account the high velocity of the UEs for bringing strong QoS support. However this task is very challenging since the high velocity of these UEs can affect the resources scheduling process and lead to performance degradations. Indeed, in LTE most of schedulers exploit the channel state in terms of CQI reports from UEs to perform the scheduling process. Moreover with the faster changing channel conditions due to the high speed environment, the channel measurements for channel estimation are not able to adapt fast enough to channel variation and the current channel conditions at the

actual scheduling instant significantly deviates from the reported channel quality. As consequences the CQI reports become outdated sooner and more and more unreliable. These erroneous and outdated feedbacks provided affect the scheduling metric and result in errors in scheduling decisions and hence degradation of the users performances. This was confirmed with our simulations where we noticed a performance degradation with the increase of the speed. Therefore these scenarios can lead to provide weak QoS support to the multimedia applications such as video traffics which require the strongest one.

To deal with this issue, we proposed two efficient strategies. The first one is a technique which maintains the required level of QoS for supporting video users at high velocities. The key idea consists in identifying depending on the UEs velocity, the minimum CQI reports rate in order to maintain the required QoS for supporting the users video flows. Simulations results showed that for a target level of PLR to maintain, less CQI reports were needed for low mobility users whereas more were required for the high mobility UEs. As for the second strategy, it is an opportunistic scheduling method which improves the performance of high speed video users. The scheme is the modified version of our E-FLS algorithm and takes into account the users velocity. With this method, the scheduler gives more priority and resources to the UEs having the highest velocity. Simulations results demonstrated its effectiveness and showed that it outperformed the E-FLS algorithm. Performance of the video users having the highest velocity have been improved since the new scheme provided the lowest values for the PLR and for the average delay of their flows.

Chapter 6

Conclusion & Future directions

In the next two sections we present the global conclusions that we have drawn from this work and the possible directions that can be investigated in the future.

6.1 Conclusion

LTE represents an emerging technology which has been designed to face the growing demand for network services by providing higher data rates, low latency and improved spectral efficiency. It is a very promising technology with respect to previous generation of mobile networks, as it is based on an all-IP architecture that aims at supporting several high quality services such as videotelephony, VoIP, video streaming and online gaming. In order to fulfill these ambitious goals, LTE introduced new technical principles and concepts. They encompass in addition to the new network architecture, the use of a new QoS management policy, the use of new multiple access schemes, and a series of mechanisms at different layers of the protocol stack, able to efficiently exploit the wireless link capacity up to the Shannon limit.

The scheduling process plays a major role in LTE networks since it is responsible for resources allocation. Indeed, an effective allocation helps to meet the system performance targets and satisfies the QoS requirements of the services. It consists in designing efficient algorithms which take into account both the QoS requirements and the physical constraints for distributing the RBs among users or flows. This process represents a great challenge since the strategy to propose must be, not only a simple and reliable algorithm which achieves a trade-off between several aspects such as fairness, overall cell throughput maximization and QoS requirements, but also a scheme which deals with the fact related to the contiguous RBs allocation required by SC-FDMA.

In this work we designed and implemented several efficient scheduling algorithms to improve the QoS of real time applications in high mobility scenarios of LTE networks. We focused on the high mobility scenarios and particularly the vehicular scenarios since they represent a significant challenge for networks operators, given the combination of their large data volumes and especially their elevate speed. Indeed for the vehicular scenarios nodes, the network could not be able to provide strong QoS support to their real time applications since their high velocity affect the resources scheduling process and lead to performance degradations. By conducting this thesis, we aimed at improving resources allocation in such scenarios by designing appropriate strategies and methods for bringing strong QoS support. To reach this goal, the study has been carried out in two steps. At first, in order to have an expert knowledge of the key facets of LTE scheduling, we conducted the study in a context where the high mobility aspect of the node was not taken into account. This helped not only to critically analyze the existing works but also to propose new solutions to improve QoS of real time applications. After that, the high mobility parameter has been added and innovative methods dealing with this context have been designed. Nevertheless due to the existing differences between the downlink and the uplink, the issue was tackled in each of the aforementioned directions.

The first contribution in Chapter 3 addressed the problem of improving the scheduling of downlink communications in a context where the high mobility was not taken into account. We designed two main schemes for this issue. The first one is an innovative strategy which improves resources assignment, particularly in the case of overbooking scenarios where we have a great number of UEs for very limited RBs. It is a scheduler which allocates RBs using the channel conditions of the users and the bearer priority of their flows. It uses a metric based on CQI reports and bearers priority for taking decisions. Fairness is enabled by using a time window and spectral efficiency is improved by scheduling twice in this time window the UEs with the best metric. Simulations have confirmed that the method allowed an acceptable level of fairness while improving the overall cell throughput in such scenarios. Due to the fact that this strategy was not a complete QoS-aware method and could not be able to guarantee the meeting of the requirements of QoS-constrained flows like RT flows, we designed a second one for this purpose. The new proposed scheme *E-FLS* is an enhanced scheduling scheme which provides strict delay bounds and guarantees very low packet loss rate to multimedia services. Its highest level works at the beginning of each frame by defining the amount of data to be transmitted in order to satisfy the required delay. As for its lowest level scheduler, it is QoS-aware and allocates RBs in each TTI in order to guarantee very low PLR to the delay-constrained flows. Simulation results confirmed that our proposed approach provided the best QoS support compared to other schemes in the literature, and increased also multimedia services performances, especially for video flows.

The second contribution in Chapter 4 concerned the problem of improving the scheduling of uplink communications in a context where the high mobility was not taken into account. We have seen that in this direction the handshake procedure consisting of a scheduling request message from the UE and a scheduling grant from the eNB required twice a communication over the air interface and caused notable delay. We have demonstrated that this delay could exceed 12 ms for the case of videotelephony flows. As consequences, the resources allocation for these types of traffics could lead to weak QoS support. We thus designed and implemented a novel scheduling protocol *SPS-P* which improves uplink resources allocation for these flows and reduces the delay caused by dynamic scheduling. The key idea consists in scheduling such flows using a semi-persistent method associated with a provisioning process. Actually, videotelephony flows are characterized in the network by variable sized packets with constant inter-arrival time. With the semi-persistent mode, the UEs do not request anymore RBs for the video frames and the provisioning mode, by using an accurate traffic prediction model, helps to define the amount of resources to reserve for each frame. A supervised learning method, namely the *SVM*, has been used to learn from a dataset of various videotelephony trace files in order to elaborate an effective traffic prediction model. We have seen afterwards that the obtained traffic prediction model was able to provide the best performance compared to the previous GBAR(1) model. Finally the performance of the proposed *SPS-P* strategy have been evaluated by simulations and the results demonstrated its effectiveness compared to the dynamic scheduling, showing that it improved videotelephony flows performance by providing the lowest value for their packet delays and by strongly reducing their packet loss ratio.

The last contribution of our thesis addressed in Chapter 5 the problem of resources allocation in high mobility scenarios. In this Chapter, the high mobility aspect was taken into account for designing suitable resources allocation schemes for vehicular scenarios. Firstly to identify the needs, we conducted a study on the impacts of the effects involved with the vehicular scenarios. This study revealed a performance degradation with the increase of the speed. Indeed, we noticed that the users with the highest velocity were the most affected and underwent significant performance degradation within all simulations compared to the other users. Clearly, vehicular scenarios could not be suitable for multimedia applications such as video flows which require strong QoS support. To cope with this issue, we proposed two efficient strategies. The first one is a technique which maintains the required level of QoS for supporting video users at high velocities. It consists in identifying depending on the UEs velocity, the minimum CQI reports rate in order to maintain the required QoS for supporting the users video flows. Simulations results showed that for a target level of PLR to maintain, less CQI reports were needed for low mobility users whereas more were required for the high mobility

UEs. The second proposed strategy is an opportunistic scheduling method which improves the performance of high speed video users. The scheme is the modified version of our *E-FLS* algorithm and takes into account the users velocity. With the proposed strategy, more priority and resources are given to the UEs having the highest velocity. Simulations results demonstrated its effectiveness and showed that it outperformed the E-FLS algorithm. Performance of the video users having the highest velocity have been improved since the new proposed scheme provided the lowest values for the PLR and for the average delay of their flows.

From this thesis we learnt that there was not a scheduling strategy adapted to all the situations. In other words, a robust scheme able to work in very different scenarios does not exist. Indeed, depending on the situation and on the goals to reach, the scheduling strategy needed to be modified and redesigned. In next section we conclude our work by describing the future directions and challenges that could be investigated.

6.2 Future directions and challenges

We present in this section the list of works that could be done in order to complete our study and the future challenges that could arise.

Regarding the additional studies, a possible work would be to improve the traffic prediction model of the proposed SPS-P mechanism of Chapter 4. Indeed, the proposed traffic prediction model has been designed only for videotelephony flows. It could be enhanced so that it considers various types of video flows. Another possible extension would be to optimize the resources allocation process itself by finding the appropriate weights to assign to the different parameters of the computed scheduling metric. Indeed when computing the scheduling metric for taking decisions, several parameters such as the channel conditions, packets delays and queue length size are taken into account and given the same importance. However since they are heterogeneous, they cannot be treated as equals. Pondering parameters could be studied in order to differentiate the importance of each parameter. Finally in this work we did not focus on the MIMO features which can be used to increase the UEs throughput. With MIMO techniques, both the eNB and the UE adopt multiple antennas for providing simultaneous transmissions of several data streams on a single radio link [86]. RBs allocation could be different when taking account this parameter since some UE could be excluded of the process. Thus additional works taking account the MIMO techniques could be interesting to carry out.

As for the future challenges, it would be essentially to adapt the proposed resources scheduling strategies for the evolution of LTE networks. Indeed, due to several new aspects introduced by the 3GPP in the release 10 for the LTE-Advanced (LTE-A) networks [87], the resources allocation strategies need to be modified and adapted. We describe below how these new parameters can influence the design of such schemes.

- 1) Scheduling with Carrier Aggregation :** Carrier Aggregation (CA) is a innovative technique introduced in the 3GPP release 10 for LTE-A in order to increase considerably the network capacity. With CA, the eNB can use up to 5 adjacent channels of 20 MHz to deliver a peak data rate of 1 Gbps and 500 Mbps for the downlink and uplink, respectively [88]. On the one hand this technique will oblige to design resources allocation strategies with *lower computational complexity* since there will be additional complexity needs. Indeed with CA, the number of RBs could reach 500 units. Considering the fact that the resources allocation policies are based on metrics computation, the scheduler could have a huge amount of metrics to compute every TTI. On the other hand, due to the fact that LTE and LTE-A will be compatible, the resources allocation should be modified in order to take into account both LTE and LTE-A users. Indeed, since LTE users cannot communicate using more than 20 MHz, the backward compatibility capabilities of LTE-A with LTE imposes to know carrier that they should use for receiving and sending their packets.
- 2) Scheduling with Multi-user MIMO :** As explained previously, MIMO techniques aims at increasing a single user throughput by exploiting the spatial multiplexing gain. However in LTE-A, MIMO can be used in a multi-user fashion (MU-MIMO). It consists in serving different UEs on different spatial streams on the same time/frequency resource, thus allocating the same RB to different UEs. This will lead also to modify the resources allocation policies. Indeed the scheduling strategies could be modified to add new decisions such as the identification of the best MU-MIMO user pair in terms of throughput maximization instead of selecting directly a single user transmission [89].
- 3) Scheduling with Coordinated Multi-Point :** Coordinated Multi-Point (COMP) transmission is a technique used in LTE-A to increase downlink throughput of a cell edge user by coordinating the downlink transmission towards this user using multiple eNBs [90]. This technique will also lead to revise the resource allocations schemes because with COMP, the scheduling process will need coordination and synchronization between different eNBs. Clearly the RBs allocation will need to be done in a distributed approach in terms of both information gathering and scheduling decisions.

4) Scheduling with Relay Nodes : LTE-A introduces also the use of very small eNBs called Relay Nodes (RN), which are designed to serve a very limited number of users within restricted coverage areas [91]. In addition to forward packets to the UEs, RNs can also be in charge of resources allocations. In these cases, communications between them, the donor eNB and the UEs are organized into a dedicated Multicast Broadcast Single-Frequency Network (MBSFN) frame structure. Concretely, RNs use particular subframes for communications with the donor eNB and leave the remaining ones for the users data transmissions. This will lead also to revise the resources allocations strategies so that they cope with the delay caused by communications with the donor eNB.

Appendix A

LTE-Sim description

Thanks to LTE-Sim which was so helpful during this thesis. Indeed, it was so difficult to find an open source framework like LTE-Sim able to efficiently simulate the LTE networks.

LTE-Sim has been designed as an event-driven simulator and written in C++ by Giuseppe Piro and Francesco Capozzi at *Politecnico di Bari*. During the thesis, we used the release 5 of the software for performing the simulations. It is composed by several classes, methods and variables which the most important are depicted in Figure A.1. More details can be found in [48] and online on www.telematics.poliba.it/index.php/en/lte-sim

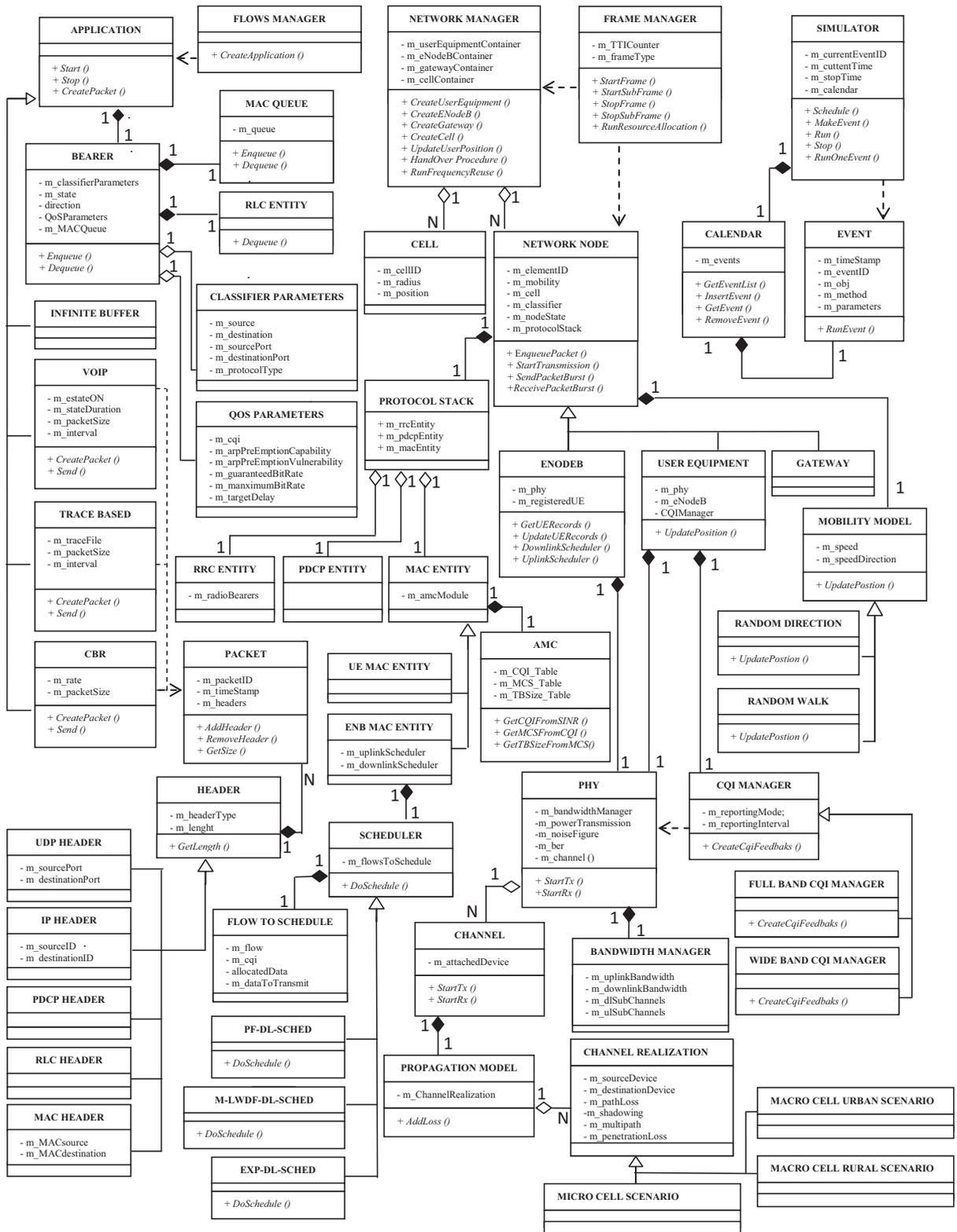


FIGURE A.1: Class diagram of LTE-Sim

Appendix B

OTDOA positioning description

OTDOA (Observed Time Difference Of Arrival) is an accurate positioning technique introduced in LTE networks for determining the UEs positions. With this method, the UE measures the time difference between the Positioning Reference Signal (PRS) from several eNBs and reports these time differences to the E-SMLC. Based on these time differences and on the knowledge of the different eNBs positions, the E-SMLC computes the final UEs location along with their speed. The process can be described by the steps below :

- 1) : The E-SMLC sends firstly through the LTE Positioning Protocol (LPP) layer an OTDOA measurements request to the UE. This request contains also all the necessary information (list of eNBs along with their PRS parameters) to help the UEs to carry out a set of Reference Signal Time Difference (RSTD) measurements.
- 2) : The UE then performs the asked measurements which consist as shown in Figure [B.1](#) in estimating the exact time offsets between the PRS from different cells. Once this task accomplished, it reports these estimated time differences along with an estimate of the measurement quality to the E-SMLC.
- 3) : Finally the E-SMLC uses the received time difference estimates and the knowledge of the eNBs location as well as transmit time offsets to compute the UE position.

More details of the procedure can be found in [\[10\]](#) and [\[92\]](#).

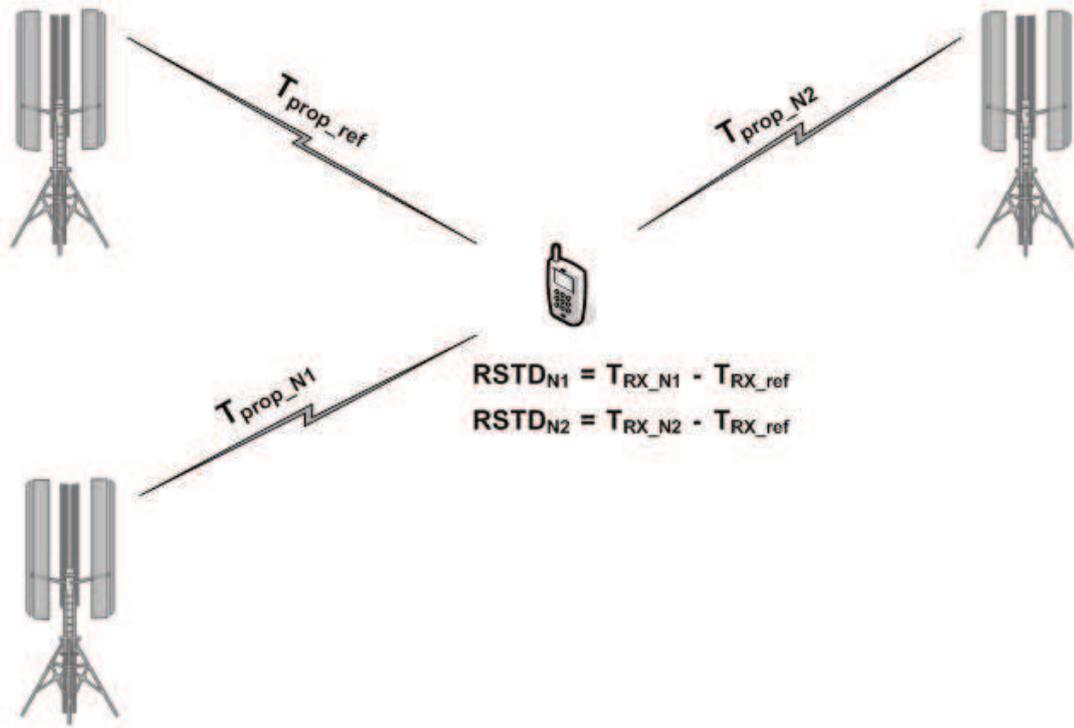


FIGURE B.1: Illustration of OTDOA in LTE

Bibliography

- [1] Harri Holma and Antti Toskala. *LTE for UMTS-OFDMA and SC-FDMA based radio access*. John Wiley & Sons, 2009.
- [2] 3GPP. Technical specifications group radio access network – requirements for evolved utra (e-utra) and evolved utran (e-utran) (release 8). *3GPP TS 25.913*, 2009.
- [3] Long Term Evolution Motorola. A technical overview. *Technical White Paper*, 2007.
- [4] Giuseppe Piro, Luigi Alfredo Grieco, Gennaro Boggia, and Pietro Camarda. A two-level scheduling algorithm for qos support in the downlink of lte cellular networks. In *2010 European Wireless Conference (EW)*, pages 246–253. IEEE, 2010.
- [5] International Telecommunication Union (ITU). Operation, overall network and service, telephone and factors, service operation and human. *ITU-T Recommendation E.800 Annex B*, August 2008.
- [6] Sondes Kallel Khemiri. *Dimensionnement et allocation de ressources dans les réseaux métropolitains IEEE 802.16*. PhD thesis, 2010.
- [7] CISCO. Cisco visual networking index: Global mobile data traffic forecast update, 2010-2015. *CISCO White Paper*, February 1st, 2011.
- [8] Abdel Mehzen Ahmad. *Techniques de transmission et d'accès sans fil dans les réseaux ad-hoc véhiculaires (VANETS)*. PhD thesis, Evry, Institut national des télécommunications, 2012.
- [9] Tayebbeh Gholipur, Mehdi Mahdavi, and Abolghasem Zeidaabadi Nezhad. A novel scheduling algorithm for low and high mobility users in multi-hop wimax network. *Wireless personal communications*, 72(2):1203–1228, 2013.
- [10] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE: the UMTS long term evolution*. Wiley Online Library, 2009.

-
- [11] Francesco Capozzi, Giuseppe Piro, Luigi Alfredo Grieco, Gennaro Boggia, and Pietro Camarda. Downlink packet scheduling in lte cellular networks: Key design issues and a survey. *Communications Surveys & Tutorials, IEEE*, 15(2):678–700, 2013.
- [12] Erik Dahlman, Stefan Parkvall, Johan Skold, and Per Beming. *3G evolution: HSPA and LTE for mobile broadband*. Academic press, 2010.
- [13] Hannes Ekstrom. Qos control in the 3gpp evolved packet system. *Communications Magazine, IEEE*, 47(2):76–83, 2009.
- [14] 3GPP. Services, technical specification group and system aspects – policy and charging control architecture (release 9). *3GPP TS 23.203*, 2011 - 2012.
- [15] 3GPP. Technical specifications group radio access network – evolved terrestrial radio access (e-utra) and evolved terrestrial radio access network (e-utran); radio resource control (rrc) protocol specification (release 9). *3GPP TS 36.331*, .
- [16] 3GPP. Technical specifications group radio access network – evolved terrestrial radio access (e-utra) and evolved terrestrial radio access network (e-utran); packet data convergence protocol (pdcp) protocol specification (release 9). *3GPP TS 36.323*, .
- [17] 3GPP. Technical specifications group radio access network – evolved terrestrial radio access (e-utra) and evolved terrestrial radio access network (e-utran); radio link control layer (rlc) protocol specification (release 9). *3GPP TS 36.322*, .
- [18] 3GPP. Technical specifications group radio access network – evolved terrestrial radio access (e-utra) and evolved terrestrial radio access network (e-utran); medium access control (mac) protocol specification (release 9). *3GPP TS 36.321*, .
- [19] 3GPP. Technical specifications group radio access network – physical channel and modulation (release 8). *3GPP TS 36.211*, 2009.
- [20] Ching-Hsiang Chuang and Phone Lin. Performance study for harq–arq interaction of lte. *Wireless Communications and Mobile Computing*, 10(11):1459–1469, 2010.
- [21] Jean Thierry Stephen Avocanh, Marwen Abdennebi, and Jalel Ben-Othman. An enhanced two level scheduler to increase multimedia services performance in lte networks. In *2014 IEEE International Conference on Communications (ICC)*, pages 2351–2356. IEEE, 2014.
- [22] Davinder Singh and Preeti Singh. Radio resource scheduling in 3gpp lte: A review. *International Journal of Engineering Trends and Technology (IJETT)*, 4(6):2405–2411, 2013.

- [23] Mauricio Iturralde, Steven Martin, and Tara Ali Yahiya. Resource allocation by pondering parameters for uplink system in lte networks. In *2013 IEEE 38th Conference on Local Computer Networks (LCN)*, pages 747–750. IEEE, 2013.
- [24] Xinyun Li. Lte uplink scheduling in multi-core systems. Master’s thesis, 2012.
- [25] Bin Liu, Hui Tian, and Lingling Xu. An efficient downlink packet scheduling algorithm for real time traffics in lte systems. In *2013 IEEE Consumer communications and networking conference (CCNC)*,, pages 364–369. IEEE, 2013.
- [26] Andrew S Tanenbaum. Modern operating systems. 2009.
- [27] Petteri Kela, Jani Puttonen, Niko Kolehmainen, Tapani Ristaniemi, Tero Henttonen, and Martti Moision. Dynamic packet scheduling performance in ultra long term evolution downlink. In *3rd International Symposium on Wireless Pervasive Computing, 2008. ISWPC 2008*, pages 308–313. IEEE, 2008.
- [28] Deming Liu and Yann-Hang Lee. An efficient scheduling discipline for packet switching networks using earliest deadline first round robin. *Telecommunication Systems*, 28(3-4):453–474, 2005.
- [29] Huda Adibah Mohd Ramli, Riyaj Basukala, Kumbesan Sandrasegaran, and Rachod Patachaianand. Performance of well known packet scheduling algorithms in the downlink 3gpp lte system. In *2009 IEEE 9th Malaysia International Conference on Communications (MICC)*,, pages 815–820. IEEE, 2009.
- [30] Mohammad T Kawser, Hasib MAB Farid, Abduhu R Hasin, Adil MJ Sadik, and Ibrahim K Razu. Performance comparison between round robin and proportional fair scheduling methods for lte. *International Journal of Information and Electronics Engineering*, 2(5):678–681, 2012.
- [31] Haidar Safa and Kamal Tohme. Lte uplink scheduling algorithms: Performance and challenges. In *2012 19th International Conference on Telecommunications (ICT)*, pages 1–6. IEEE, 2012.
- [32] Fang Liu, Xiaoming She, Lan Chen, and Hiroyuki Otsuka. Improved recursive maximum expansion scheduling algorithms for uplink single carrier fdma system. In *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, pages 1–5. IEEE, 2010.
- [33] Riyaj Basukala, Huda Adibah Mohd Ramli, and Kumbesan Sandrasegaran. Performance analysis of exp/pf and m-lwdf in downlink 3gpp lte system. In *First Asian Himalayas International Conference on Internet, 2009. AH-ICI 2009*, pages 1–5. IEEE, 2009.

- [34] Sadiq Bilal, Madan Ritesh, and Sampath Ashwin. Downlink scheduling for multi-class traffic in lte. *EURASIP Journal on Wireless Communications and Networking*, 2009, 2009.
- [35] Daniela Laselva, Francesco Capozzi, Frank Frederiksen, Klaus I Pedersen, Jeroen Wigard, and István Z Kovács. On the impact of realistic control channel constraints on qos provisioning in utran lte. In *2009 IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, pages 1–5. IEEE, 2009.
- [36] 3GPP. Technical specifications group radio access network – persistent scheduling in e-utra. *3GPP TSG-RAN WG1 R1-070098*, .
- [37] Dajie Jiang, Haiming Wang, Esa Malkamaki, and Esa Tuomaala. Principle and performance of semi-persistent scheduling for voip in lte system. In *International Conference on Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007*, pages 2861–2864. IEEE, 2007.
- [38] Yong Fan, Petteri Lundén, Markku Kuusela, and Mikko Valkama. Efficient semi-persistent scheduling for voip on eutra downlink. In *IEEE 68th Vehicular Technology Conference, 2008. VTC 2008-Fall*, pages 1–5. IEEE, 2008.
- [39] Liljana Gavrilovska and Dejan Talevski. Novel scheduling algorithms for lte downlink transmission. In *2011 19th Telecommunications Forum (TELFOR)*, pages 398–401. IEEE, 2011.
- [40] Jean Thierry Stephen Avocanh, Marwen Abdennebi, and Jalel Ben-Othman. A new two-level scheduling algorithm for the downlink of lte networks. In *2013 IEEE Globecom Workshops (GC Wkshps)*, pages 4519–4523. IEEE, 2013.
- [41] Akhilesh Pokhariyal, Klaus I Pedersen, Guillaume Monghal, Istvan Z Kovacs, Claudio Rosa, Troels E Kolding, and Preben E Mogensen. Harq aware frequency domain packet scheduler with different degrees of fairness for the utran long term evolution. In *VTC2007-Spring. IEEE 65th Vehicular Technology Conference, 2007*, pages 2761–2765. IEEE, 2007.
- [42] Xavier Gelabert, Jordi Pérez-Romero, Oriol Sallent, and Ramón Agusti. A 4-dimensional markov model for the evaluation of radio access technology selection strategies in multiservice scenarios. In *2006 IEEE 64th Vehicular Technology Conference, 2006. VTC-2006 Fall*, pages 1–5. IEEE, 2006.
- [43] Christian Mehlführer, Martin Wrulich, J Colom Ikuno, Dagmar Bosanska, and Markus Rupp. Simulating the long term evolution physical layer. In *Proc. of the 17th European Signal Processing Conference (EUSIPCO 2009), Glasgow, Scotland*, volume 27, page 124. Citeseer, 2009.

- [44] Raj Jain, Dah-Ming Chiu, and William Hawe. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. 1998.
- [45] Richard Musabe and Hadi Larijani. Evaluation of a new scheduling scheme for voip with mobility in 3g lte. In *CTRQ 2012, The Fifth International Conference on Communication Theory, Reliability, and Quality of Service*, pages 7–13, 2012.
- [46] Sunggu Choi, Kyungkoo Jun, Yeonseung Shin, Seokhoon Kang, and Byoungjo Choi. Mac scheduling scheme for voip traffic service in 3g lte. In *2007 IEEE 66th Vehicular Technology Conference, 2007. VTC-2007 Fall*, pages 1441–1445. IEEE, 2007.
- [47] Karl J Aström and Bjorn Wittenmark. *Computer-controlled systems: theory and design*. Courier Corporation, 2011.
- [48] Giuseppe Piro, Luigi Alfredo Grieco, Gennaro Boggia, Francesco Capozzi, and Pietro Camarda. Simulating lte cellular systems: an open-source framework. *IEEE Transactions on Vehicular Technology*, 60(2):498–513, 2011.
- [49] Video trace library. <http://trace.eas.asu.edu>.
- [50] Chen-Nee Chuah and Randy H Katz. Characterizing packet audio streams from internet multimedia applications. In *IEEE International Conference on Communications, 2002. ICC 2002*, volume 2, pages 1199–1203. IEEE, 2002.
- [51] Stefan Schwarz, Christian Mehlhfuhrer, and Markus Rupp. Low complexity approximate maximum throughput scheduling for lte. In *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1563–1569. IEEE, 2010.
- [52] Markus Laner, Philipp Svoboda, Peter Romirer-Maierhofer, Navid Nikaein, Fabio Ricciato, and Markus Rupp. A comparison between one-way delays in operating hspa and lte networks. In *2012 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pages 286–292. IEEE, 2012.
- [53] A Alheraish, S Alshebeili, and T Alamri. Regression video traffic models in broadband networks. *Journal of King Saud University*, 2005.
- [54] Savera Tanwir and Harry G Perros. A survey of vbr video traffic models. *IEEE Communications Surveys and Tutorials*, 15(4):1778–1802, 2013.
- [55] Daniel P Heyman, Ali Tabatabai, and TV Lakshman. Statistical analysis and simulation study of video teleconference traffic in atm networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 2(1):49–59, 1992.

- [56] Yuzhe Xu. Latency and bandwidth analysis of lte for a smart grid. 2011.
- [57] Xueming Pan, Guoqing Li, and Yanan Lin. Method and user equipment for sending uplink resource scheduling request, September 15 2010. US Patent App. 13/498,682.
- [58] Hongkun Yang, Fengyuan Ren, Chuang Lin, and Jiao Zhang. Frequency-domain packet scheduling for 3gpp lte uplink. In *2010 Proceedings IEEE INFOCOM*, pages 1–9. IEEE, 2010.
- [59] Francesco Davide Calabrese, Per-Henrik Michaelsen, Claudio Rosa, Mohmmad Anas, Carlos Ubeda Castellanos, Dimas López Villa, Klaus I Pedersen, and Preben E Mogensen. Search-tree based uplink channel aware packet scheduling for utran lte. In *VTC Spring 2008. IEEE Vehicular Technology Conference, 2008.*, pages 1949–1953. IEEE, 2008.
- [60] Akash Baid, Ritesh Madan, and Ashwin Sampath. Delay estimation and fast iterative scheduling policies for lte uplink. In *2012 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pages 89–96. IEEE, 2012.
- [61] Daniel P Heyman. The gbar source model for vbr videoconferences. *IEEE/ACM Transactions on Networking (TON)*, 5(4):554–560, 1997.
- [62] TV Lakshman, Partho Pratim Mishra, and KK Ramakrishnan. Transporting compressed video over atm networks with explicit-rate feedback control. *IEEE/ACM Transactions on Networking (TON)*, 7(5):710–723, 1999.
- [63] Jean Thierry Stephen Avocanh, Marwen Abdennebi, Jalel Ben-Othman, and Giuseppe Piro. A semi-persistent scheduling scheme for videotelephony traffics in the uplink of lte networks. In *Proceedings of the 17th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, pages 321–325. ACM, 2014.
- [64] Thorsten Joachims. *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.
- [65] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.
- [66] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [67] Vladimir Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2000.

- [68] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. In *The Journal of Machine Learning Research*, pages 281–305, 2012.
- [69] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145, 1995.
- [70] Avishek Patra, Volker Pauli, and Yu Lang. Packet scheduling for real-time communication over lte systems. In *2013 IFIP Wireless Days (WD)*, pages 1–6. IEEE, 2013.
- [71] Xiaoyue WANG and Di HE. Performance analysis of lte downlink system with high velocity users. *Journal of Computational Information Systems*, 10(9):3645–3652, 2014.
- [72] The Free Library. *Computers in railways XIII*. Computer system design and operation in railways and other transit systems, 2012.
- [73] Guillaume Rémy, S-M Senouci, François Jan, and Yvon Gourhant. Lte4v2x-impact of high mobility in highway scenarios. In *2011 Global Information Infrastructure Symposium (GIIS)*,, pages 1–7. IEEE, 2011.
- [74] Guillaume Remy, S-M Senouci, François Jan, and Yvon Gourhant. Lte4v2x: Lte for a centralized vanet organization. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–6. IEEE, 2011.
- [75] Sandesh Uppoor and Marco Fiore. Characterizing pervasive vehicular access to the cellular ran infrastructure: an urban case study.
- [76] Ralf-Peter Schafer, Stefan Lorkowski, Peter Mieth, and Ingo Schurr. Travel time measurements using gsm and gps probe data. In *16th ITS World Congress and Exhibition on Intelligent Transport Systems and Services*, 2009.
- [77] Meihui TrafficCast. <http://www.meihuichina.com/>.
- [78] Andreas Muller and Tao Chen. On the effects of different user speeds on the performance of high speed downlink packet access. In *Proc. 14th IST Mobile and Wireless Communications Summit*, 2005.
- [79] Andras Racz, Andras Temesvary, and Norbert Reider. Handover performance in 3gpp long term evolution (lte) systems. In *16th IST Mobile and Wireless Communications Summit, 2007.*, pages 1–5. IEEE, 2007.
- [80] Mohmmad Anas, Francesco Davide Calabrese, P-E Ostling, Klaus I Pedersen, and Preben E Mogensen. Performance analysis of handover measurements and layer 3 filtering for utran lte. In *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007.*, pages 1–5. IEEE, 2007.

- [81] Konstantinos Dimou, Min Wang, Yu Yang, Muhammad Kazmi, Anna Larmo, Jonas Pettersson, Walter Muller, and Ylva Timmer. Handover within 3gpp lte: design principles and performance. In *2009 IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*,, pages 1–5. IEEE, 2009.
- [82] D Morales-Jiménez, Juan J Sánchez, G Gómez, MC Aguayo-Torres, and JT En-trambasaguas. Impact of feedback channel delay on adaptive ofdma systems. In *ERCIM Workshop on eMobility, Coimbra, Portugal*, pages 1–10, 2007.
- [83] Stanislav Nonchev, Mikko Valkama, and Ridha Hamila. Effect of high-velocity scenarios on the performance of mimo lte packet scheduling. In *2011 8th international multi-conference on Systems, signals and devices (SSD)*,, pages 1–6. IEEE, 2011.
- [84] Riyaj Basukala, HAM Ramli, Kumbesan Sandrasegaran, and Lu Chen. Impact of cqi feedback rate/delay on scheduling video streaming services in lte downlink. In *2010 12th IEEE International Conference on Communication Technology (ICCT)*,, pages 1349–1352. IEEE, 2010.
- [85] Mohamed Nidhal Mejri, Jalel Ben-Othman, and Mohamed Hamdi. Survey on vanet security challenges and possible cryptographic solutions. *Vehicular Communications*, 1(2):53–66, 2014.
- [86] Kian Chung Beh, Angela Doufexi, and Simon Armour. On the performance of su-mimo and mu-mimo in 3gpp lte downlink. In *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*,, pages 1482–1486. IEEE, 2009.
- [87] Stefan Parkvall, Anders Furuskar, and Erik Dahlman. Evolution of lte toward int-advanced. *IEEE Communications Magazine*,, 49(2):84–91, 2011.
- [88] Klaus Ingemann Pedersen, Frank Frederiksen, Claudio Rosa, Hung Nguyen, Luis Guilherme Uzeda Garcia, and Yuanye Wang. Carrier aggregation for lte-advanced: functionality and performance aspects. *IEEE Communications Magazine*,, 49(6): 89–95, 2011.
- [89] Bishwarup Mondal, Timothy A Thomas, and Amitava Ghosh. Mu-mimo system performance analysis in lte evolution. In *2010 IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*,, pages 1510–1515. IEEE, 2010.
- [90] Ralf Irmer, Heinz Droste, Patrick Marsch, Michael Grieger, Gerhard Fettweis, Stefan Brueck, H-P Mayer, Lars Thiele, and Volker Jungnickel. Coordinated multipoint: Concepts, performance, and field trial results. *IEEE Communications Magazine*,, 49(2):102–111, 2011.

-
- [91] Aleksandar Damnjanovic, Juan Montojo, Yongbin Wei, Tingfang Ji, Tao Luo, Madhavan Vajapeyam, Taesang Yoo, Osok Song, and Durga Malladi. A survey on 3gpp heterogeneous networks. *IEEE Wireless Communications*, 18(3):10–21, 2011.
- [92] 3GPP. Universal terrestrial radio access (utra) and evolved utra (e-utra) and evolved packet core (epc); user equipment (ue) conformance specification for ue positioning; part 1: Conformance test specification. *3GPP TS 37.571-1*, .