Downlink Radio Resource Management for QoS Provisioning in OFDMA Systems

with emphasis on Admission Control and Packet Scheduling

Monghal, Guillaume Damien

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Downlink Radio Resource Management for QoS Provisioning in OFDMA systems

- with Emphasis on Admission Control and Packet Scheduling

PhD Thesis

by

Guillaume Damien André Monghal

A Dissertation submitted to
the Faculty of Engineering, Science and Medicine, Aalborg University
in partial fulfillment for the degree of
Doctor of Philosophy,
Aalborg, Denmark,
May 2009.

Supervisors:
Preben Elgaard Mogensen, PhD,
Professor, Aalborg University, Denmark.
Klaus Ingemann Pedersen, PhD,
Senior Wireless Networks Specialist, Nokia Siemens Networks, Aalborg, Denmark.

Opponents:
Associate Professor Tatiana Madsen, PhD,
Aalborg University, Denmark.
Professor Lars Dittmann, PhD,
Technical University of Denmark, Denmark.
Mikko Rinne, Msc,
Senior Research Manager, Nokia, Finland.

ISBN: 978-87-92328-14-4

Abstract

Orthogonal Frequency Division Multiple Access (OFDMA) is the preferred technology for future downlink mobile broadband access systems as the 3rd Generation Partnership Project Long Term Evolution (LTE) in downlink, where the diversification of the proposed services (Voice Over Internet Protocol, video streaming, gaming or simple web browsing...) and higher throughputs are key targets. Therefore, Quality of Service in OFDMA is a key issue for the success of next generation mobile systems. The present thesis aims at proposing a concrete Quality of Service aware packet scheduling and Radio Admission Control solution for a realistic OFDMA based system in downlink where LTE is taken as a case study.

In the frame of the thesis work, a detailed system model of LTE dedicated to Radio Resource Management study has been developed and implemented in a semi-static system level simulator.

In a first phase, the study focuses on the dual time / frequency domain packet scheduling concept which allows a linear complexity. Fundamental design principles for throughput control are highlighted. Using those principles, the design of a complete Quality of Service aware packet scheduling algorithm based on the Required Activity Detection principle is proposed. The time domain Required Activity Detection algorithm shows to be among the best performing algorithms in terms of Quality of Service outage. However, the study shows that it can be greatly improved by introducing delay awareness. In the frequency domain, Required Activity Detection applies weights to the frequency domain schedulers and when used, can decrease the outage of up to 50%.

Packet scheduling also is studied in fractional load condition defined by a partial use of the frequency domain resource. The study shows that, fractional load cannot be used as a coverage enhancement technique with the packet scheduling algorithms of the thesis, furthermore, when fractional load occurs due to a lack of traffic, the block error rate can increase severely thus affecting the experienced Quality of Service. The thesis proposes various solutions to overcome that problem, among which the simplest consists in using Wideband Interference Reporting in the Channel Quality Information reporting scheme.

Finally, the thesis focuses on Radio Admission Control. New algorithms, namely the adaptive throughput and Required Activity Detection-based Radio Admission Control are tested and compared to a simple fixed throughput algorithm. It is shown that the Required Activity Detection based Radio Admission Control can track the channel and therefore keep the system in feasibility region.

Dansk Resumé [1]

Orthogonal Frequency Division Multiple Access (OFDMA) er den foretrukne teknologi for mobile bredbåndssystemer såsom 3GPP LTE, Downlink. Den vigtigste kommercielle interesse i sådanne systemer er en differentiering mellem services (VoIP, video streaming, spil eller almindelig web browsing) ved høje systembelastninger. I denne afhandling studeres forskellige løsninger til packet scheduling (PS) og radio adgangskontrol løsninger til et realistisk system baseret på OFDMA. Indenfor disse rammer præsenteres en detaljeret systemmodel af LTE med fokus på RRM. Denne model er implementeret i en semi-statisk system simulator.

I første fase fokuserer studiet på PS strategier med lav kompleksitet og rimelig fordeling af data- hastighed mellem brugerne. Adskillige PS algoritmer foreslås baseret på en todelt tids-frekvens PS løsning med lav kompleksitet. To fundamentale design principper fremhæves i løbet af studiet. For det første, uafhængigheden mellem tids- and frekvens-PS for at sikre deres stabilitet og den ønskede konvergens. For det andet skal kontrol af datahastighed til de enkelte brugere udelukkende håndteres af tidsdomæne PS, når der er et stort antal brugere, og af frekvensdomæne PS, når der er et lille antal brugere.

PS algoritmen udvikles videre baseret på “Required Activity Detection (RAD)” princippet. Tidsdomæne RAD PS algoritmen viser sig at være en af de algoritmer, der har den bedste performance, når bestemte kvalitetskrav som f.eks. datahastighed skal opfyldes for de enkelte brugere. Imidlertid viser studiet, at det kan forbedres meget ved også at introducere pakke delay afhængighed i PS algoritmen. Benyttes RAD princippet også i frekvensdomænet, opnås der yderligere forbedringer.

Specielle PS algoritmer optimeret til lav load, hvor kun en del af system båndbredden bruges, er også en del af studiet. Her vises det, at ved lav load stiger “Block Error Rate (BLER)” ofte væsentligt, og vil således have indflydelse på servicekvaliteten. Der foreslås forskellige løsninger på det problem, hvor den mest enkle består i at bruge bredbånds radio kvalitetsmålinger fra den enkelte bruger, og ikke frekvensselektive målinger, som normalt er at foretrække. Alternative PS algoritmer i de tilfælde, hvor brugerne rapporterer frekvensselektive målinger, udledes også samt evalueres.

Til sidst fokuseres der på radio adgangskontrol. Nye algoritmer, RAD baserede algoritmer, testes og sammenlignes med mere simple algoritmer, der antager konstant kapacitet per celle. Det demonstreres, at RAD baseret adgangskontrol tager kanal kvaliteten for den nye bruger i betragtning, før der tages beslutninger. Algoritmen kontrollerer systembelastningen, så alle brugere i systemet kan få opfyldt deres minimums data-hastighedskrav.

[1] Jytte Larsen & Klaus I. Pedersen, Nokia Siemens Networks, Aalborg, Denmark.

Preface and Acknowledgments

This dissertation is the result of a three years research project carried out at the Radio Access Technology (RATE) section, Institute of Electronic Systems, Aalborg University, Denmark, under the supervision and guidance of Professor Preben E. Mogensen (Aalborg University, Denmark) and Dr. Klaus I. Pedersen (Nokia Siemens Networks, Aalborg, Denmark). The dissertation has been completed in parallel with the mandatory course work, teaching, and project work obligations in order to obtain the PhD degree. This research project has been co-financed by Aalborg University and Nokia Siemens Networks R&D, Aalborg.

I am sincerely grateful to my supervisors Preben E. Mogensen and Klaus I. Pedersen. Their strong technical knowledge united with their leading skills always made feel confident regarding the completion of my PhD and for the quality of my work. And even more importantly to me, they have always shown me their human understanding, patience and support along the three years of my study. I feel very sincerely privileged, honored and thankful for having worked with such a great supervisor team.

I would like to thank my Msc. supervisor Christian Rom. Firstly for proposing me a great master project and through his inspired supervising, revealing my motivation to pursue a PhD study. Christian helped me value my strong sides and therefore helped me taking decisions that I really feel were right for me. Furthermore, I want to thank him for introducing me to the RATE group, a great place to work, and environment where I felt extremely happy on both professional and personal level.

I would also like to thank my colleagues and friends Sanjay Kumar and Akhilesh Pokhariyal who carried out their PhD project on overlapping periods and in the same research area leading to a fruitful collaboration and several common publications. I want to thank as well other researchers, with whom I had the chance to work and who were specially inspirational to me including Mohmmad Anas, Per-Henrik Michaelsen, Francesco Davide Calabrese, Jens Steiner, Mads Brix, Jeroen Wigard, Daniela Laselva, Frank Frederiksen, Dimas Lopez, Gilberto Berardinelli, Malek Boussif, Claudio Rosa, Istvan Kovacs.

It is important for me to thank Lisbeth Schiønning Larsen and Jytte Larsen who reduced my administrative life to “passing to see them from time to time”, enabling me to focus exclusively on my PhD study. I would like to also give special thanks to Lisbeth for her kindness and patience in any situation. My workplace gave me the chance to discover a new important part of me life: running. I thank Patrick for sharing this sport through which I could always find simple joy. I thank Oumer and Jeroen for talking me into running a marathon. One of the greatest and most intense experiences in my life. Furthermore, I would like to thank all the past and present colleagues at Radio Access Technology section, Aalborg University and Nokia Siemens Networks, Aalborg for their friendly support.

Finally, my warmest thanks to my dearest friends Oumer, Francesco and Anas. I am privileged to have met such friends. I am grateful to them for their unconditional support and friendship in difficult times. I am gratefully to them for sharing their passionate enthusiasm for life at any time.

Notation

Abbreviations and mathematical conventions used in the thesis are listed below for quick reference. The abbreviations are additionally defined at their first occurrence.

Abbreviations

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
ACK	Acknowledgement
AMBR	Agregate Maximum Bit Rate
AMPS	Advanced Mobile Phone Services
AOL	Average Offered Load
ARP	Allocation and Retention Priority
ARQ	Automatic Repeat ReQuest
AVI	Actual Value Interface
BE	Best Effort
BET	Blind Equal Throughput
BFF	Bandwidth Fraction Factor
BLER	BLock Error Rate
BM	Best Metric
BQPP	Best Quality Physical Resource Block (PRB) Pattern
BS	Base Station
CBR	Constant Bit Rate
CDF	Cumulative Density Function
CoItA	Carrier over Interference to Average
CQI	Channel Quality Information
CRC	Cyclic Redundancy Check
eNode-B	E-UTRAN Node B
EESM	Exponential Effective SINR Metric

EPC	Evolved Packet Core
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FDPS	Frequency-Domain Packet Scheduling
FD-RAD	Frequency-Domain Required Activity Detection
FFT	Fast Fourier Transform
FTP	File Transfer Protocol
GBR	Guaranteed Bit Rate
GPF	Generalized Proportional Fair
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
HARQ	Hybrid Automatic Repeat reQuest
HSDPA	High-Speed Downlink Packet Access
i.i.d	Independent and Identically Distributed
IFFT	Inverse Fast Fourier Transform
ICI	Inter-Carrier Interference
ICIC	Inter-Cell Interference Coordination
IEEE	Institute of Electrical and Electronis Enginners
IMS	IP Media Subsystem
IP	Internetworks Protocol
LA	Link Adaptation
LTE	Long Term Evolution
MAC	Medium Access Control
MBR	Maximum Bit Rate
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
M-LWDF	Modified Largest Weighted Delay First
MME	Mobility Management Entity
MRC	Maximal Ratio Combining
NACK	Non-Acknowledgement
N-BET	Normalized Blind Equal Throughput
NMT	Nordic Mobile Telephone
NSN	Nokia Siemens Networks
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OLLA	Outer Loop Link Adaptation
PDB	Packet Delay Budget
PDCP	Packet Data Convergence Protocol
PELR	Packet Error Loss Rate

P-GW	Public Data Network Gateway
PDCCH	Packet Downlink Control CHannel
PDSCH	Physical Downlink Shared CHannel
PDU	Protocol Data Unit
PF	Proportional Fair
PFsch	Proportional Fair scheduled
PHY	PHYsical Layer
PRB	Physical Resource Block
PS	Packet Scheduler
PSS	Priority Set Scheduler
PUCCH	Physical Uplink Control CHannel
PUSCH	Physical Uplink Shared CHannel
QCI	Quality of Service Class Identifier
QoS	Quality of Service
RAC	Radio Admission Control
RAD	Required Activity Detection
RAN	Radio Access Network
RCPP	Random Correlated PRB Pattern
RLC	Radio Link Control
RNC	Radio Network Controler
RR	Round Robin
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Reference Signal Strength Indicator
SC-FDMA	Single Carrier Frequency Division Multiple Access
SDU	Service Data Unit
SFR	Soft Frequency Reuse
S-GW	Serving Gateway
SINR	Signal-to-Interference-plus-Noise Ratio
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TDPS	Time Domain Packet Scheduling
TD-RAD	Time-Domain Required Activity Detection
TTA	Throughput To Average
TTI	Transmission Time Interval
TU	Typical Urban

UDO	User Diversity Order
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UTRAN	Universal Terrestrial Radio Access Network
VoIP	Voice Over Internet Protocol
WiMAX	Worldwide Interoperability for Microwave Access
WIR	Wideband Interference Reporting
WCDMA	Wideband Code Division Multiple Access
WRR	Weighted Round Robin

Mathematical Conventions

The following mathematical conventions are used throughout the thesis:

$\mathbf{a}, \mathbf{A}$	Bold upper or lower case indicates a vector.
$\mathbf{a}[n], \mathbf{A}[n]$	Vectors are indexed between square brackets.
a, A	Non-bold indicates a scalar.
$\lceil a \rceil$	Rounding a up to the nearest integer.
$\lvert a \rvert$	Absolute value of a.
$\overline{x}$	indicates the mean of x.
$\hat{x}$	indicates an estimator of x.
$\overline{\mathbf{a}\left[\cdot\right]_i}$	arithmetique average of the values of vector $\mathbf{a}$ indexed on the $\cdot$ (and usually indexed by i).
$\mathrm{Var}\left\{x\right\}$	Variance of random variable x, $\mathrm{Var}\left\{x\right\} = \mathrm{E}\left\{\left(x - \mathrm{E}\left\{x\right\}\right)^2\right\}$.
$\mathrm{Std}\left\{x\right\}$	Standard deviation of random variable x, $\mathrm{Std}\left\{x\right\} = \sqrt{\mathrm{Var}\left\{x\right\}}$.

Contents

Chapter 1

Thesis Introduction

In section 1.1 of this first chapter, we present the key points in the recent history of mobile communication relating to the PhD study topic in order to provide a clear understanding of the stakes. Then as the 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) is the technology case study used in our thesis, basic standardization information, essential to understanding the components of LTE at stake in PhD study, are given in section 1.2. The main principles of Orthogonal Frequency Division Multiple Access (OFDMA) as multiple access technology of LTE are briefly introduced in section 1.3. The state of the art of Radio Resource Management (RRM) for OFDMA is introduced in section 1.4. Finally, we explain the PhD study motivations and objectives in section 1.5, the general methodology in section 1.6, the novelty and contributions of the thesis in section 1.7 and the chapter outline of the thesis is described in section 1.8.

1.1 Preliminaries

In a Radio Access Network (RAN), Quality of Service (QoS) is defined as the ability of a network to provide a service to an end user at a given service level. A simple example of QoS is the ability of a mobile network to deliver an internet streaming video on a mobile device so that the video can be watched comfortably without interruption. RRM is the set of components in the radio access network that help achieving QoS while using efficiently the available transmission resource.

Mobile networks types are usually classified in terms of generations to describe the evolution of technologies and capabilities. Figure 1.1 illustrates the evolution of mobile networks. The First Generation (1G) mobile systems like the Advanced Mobile Phone Services (AMPS) or Nordic Mobile Telephone (NMT) were fist designed in the 1970's. They are analogue systems providing voice call services [1]. The Second Generation (2G) introduced digital communications in the 1980's. The Global System for Mobile Communication (GSM) became the most commercially successful 2G system as it was the first fully specified system with international compatibility and transparency [2]. Until GSM, mobile networks were purely circuit switched networks. However, in the 1990's the revolution of the Internet pushed the standardization forward. A packet switched core network was added on top of the traditional circuit switched GSM core network under the name of General Packet Radio Service (GPRS). This enables to provide basic packet based services to mobile users like Internet over the Wireless Application Protocol. In the first versions, QoS was supported in GPRS only at the core network level as the GSM radio interface

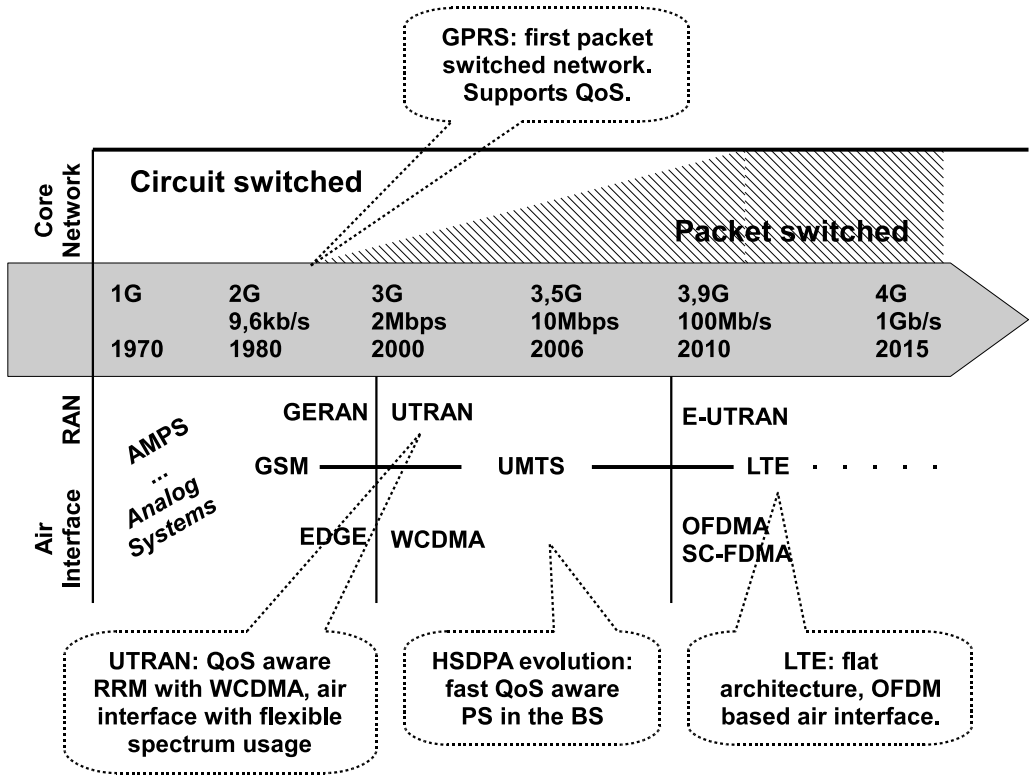

Figure 1.1: Evolution of Mobile Networks

was designed for circuit switched connections [2].

Development of Third Generation (3G) systems was steered by the increasing demand for mobile packet services and higher data rates. The international standardization body 3GPP started in 1998 the specification of the Universal Mobile Telecommunications System (UMTS) based on the existing GSM specification. A new RAN was introduced: the Universal Terrestrial Radio Access Network (UTRAN) together with a new air interface: Wideband Code Division Multiple Access (WCDMA). WCDMA allows a very flexible usage of the spectrum thanks to advanced Link Adaptation (LA) and Power Control techniques [3], [4]. The flexibility of the radio air interface gives a critical importance to QoS aware RRM. High-Speed Downlink Packet Access (HSDPA) is an enhancement brought to UTRAN which, among others, brings the Packet Scheduler (PS) and LA closer to the air interface directly in the Base Station (BS) [3]. This allows a faster adaptation to the channel and therefore more flexibility and data rates increased from 2Mbps up to 10Mbps.

Fourth Generation (4G) systems are expected to formalize the convergence between mobile networks and wireless LAN systems into "broadband wireless access" [5]. The 3GPP started the standardization of the LTE, a new all IP mobile network system also known as 3.9G for which the first products will be available in 2010. LTE prepares the way toward 4G systems with a new simplified core network and RAN architecture in order to reduce latency for packet based traffic. The new RAN called Evolved Universal Terrestrial Radio Access Network (E-UTRAN) is composed of only one node, the E-UTRAN Node B (eNode-B), which carries all the RRM functionali-

ties. Furthermore, in order to make broadband system feasible, the new downlink air interface is based on Orthogonal Frequency Division Multiplexing (OFDM). contrarily to WCDMA, OFDM based radio access techniques allow low complex receiver on bandwidths larger than 5MHz, and, OFDMA does not produce inter cell interference [3], [6].

The radio access technique for LTE downlink is OFDMA. OFDMA consists in multiplexing different users in time and frequency domain. OFDMA and LTE offer new challenges and new possibilities in terms of RRM. Firstly, as HSDPA introduced the possibility of performing fast channel aware and QoS aware PS in the time domain, OFDMA adds the frequency domain dimension. The two dimensional adaptation is a key feature for increasing the cell capacity. Furthermore, in E-UTRAN, the Radio Admission Control (RAC) functionality is placed in the eNode-B. RAC therefore benefits from a close proximity to the air interface. The thesis focuses on the design of QoS aware advanced PS and RAC for OFDMA using LTE as a study case.

Worldwide Interoperability for Microwave Access (WiMAX) radio access technology based on the 802.16 air interface standardized by the Institute of Electrical and Electronis Enginners (IEEE) starts to emerges and offers similar purpose as well as similar technical solution than LTE [7]. Most of the thesis findings can therefore be applied to WiMAX.

1.2 LTE / E-UTRAN

1.2.1 Goals and Targets

The LTE study item was launched in 2004 while the development of WCDMA was still ongoing at full speed. The aim behind LTE was to create a new mobile network system meeting the future needs of the market. The feasibility study was launched in 2005 in order to define the best radio interface and network architecture. The main requirements for E-UTRAN downlink defined in [8] are:

- Packet-Switch domain optimized; Future communications are often seen as packet based only,
- Server to User Equipment (UE) round trip time below 30ms and access delay below 300ms,
- Peak rate uplink/dowlink 50/100 Mbps,
- Good level of mobility and security,
- Improved terminal power efficiency
- Frequency allocation flexibility with 1.4, 3, 5, 10, 15 and 20 MHz allocation,
- Higher capacity compared to HSDPA reference case.

The performance studies summarized in [9] showed that the requirements were achievable with OFDMA and a flat network architecture.

OFDMA provides several advantages over WCDMA. Firstly, it is possible to implement low complexity receiver thanks to the Inverse Fast Fourier Transform (IFFT) and Fast Fourier

Transform (FFT) algorithms. Furthermore, in order to achieve the expected bit-rates, it is required for LTE to use advanced antenna techniques like Multiple Input Multiple Output (MIMO). Finally, OFDMA is a good support for frequency division multiple access techniques as it offers the possibility to send orthogonal signals to different UEs on different frequency chunks. While WCDMA would require complex equalizer due to increased multipath component on a bandwidth larger than 5MHz, OFDMA is simply free from Inter-Carrier Interference (ICI). The multiple access techniques chosen for downlink is OFDMA and for uplink is Single Carrier Frequency Division Multiple Access (SC-FDMA). As the thesis deals with downlink, we will focus only on OFDMA.

The second important technological break is the introduction of a flat architecture. The RAN E-UTRAN is composed of only one node, the eNode-B. All the radio control functionalities of the RAN are in the BS. Combined with OFDMA, this enables fast channel aware packet scheduling in both time and frequency domain. Simulations show that the cell capacity can improve up to 40% with this technique [10]. Furthermore, the flat architechture allows to lower the round trip time and packet delay in general as packet have to travel through less network nodes. However, flat architecture is possible at the cost of macro diversity gain that was managed by the Radio Network Controler (RNC) in UTRAN.

1.2.2 LTE Overall Architecture

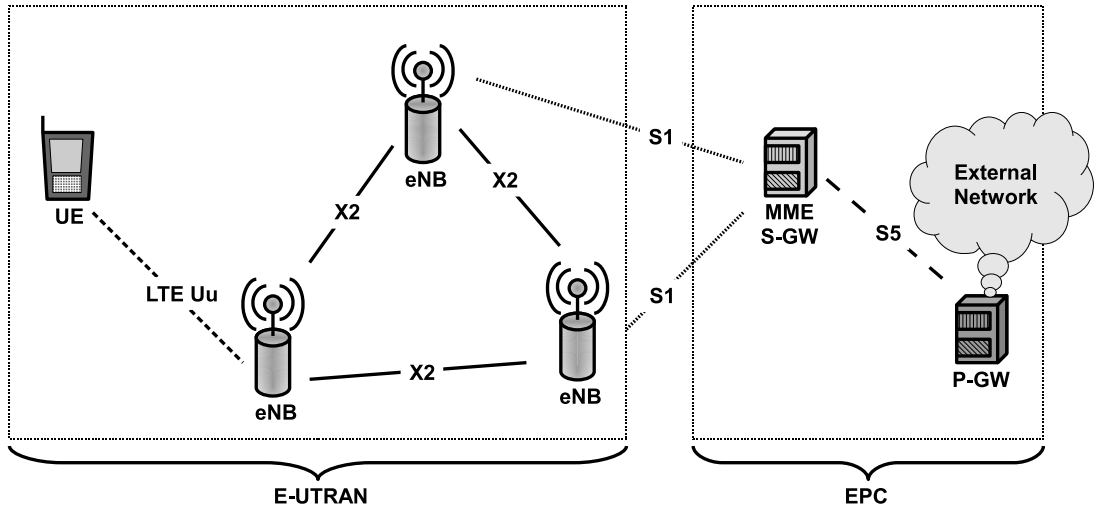

Figure 1.2: General simplified LTE architecture

The 3GPP LTE includes specifications for a core network, the Evolved Packet Core (EPC) and for a RAN, the E-UTRAN. The mains characteristics of EPC are a simplified architecture for higher throughput and lower latencies and mobility management between different types of RAN including non 3GPP (WiMAX for example). The main nodes are the Mobility Management Entity (MME), the Serving Gateway (S-GW) and the Public Data Network Gateway (P-GW). The MME is a control entity that manages the connections with the RAN and performs authentication. The S-GW forwards packets to the RAN. The P-GW is the anchor point of the UE that stays fixed throughout the connection and which is directly connected to the external network. EPC is described in the specifications [11] and [12]. As mentioned in section 1.2.1, E-UTRAN is

composed of only one node, the eNode-B which carries all the RRM functionalities as well as radio functionalities. E-UTRAN is described in the specification [13]. Figure 1.2 depicts a simple LTE architecture scenario including the main standardized interfaces.

1.2.3 Quality of Service (QoS) in LTE

As mentioned before, QoS is defined as the ability of a network to provide a service to an end user at a given service level [14]. More precisely, a Service level corresponds to a set of objective parameters named QoS parameters relating directly to end user experience, for example: packet delay or bit rate. In mobile networks, QoS encompasses all the different mechanisms that insure compliance with the QoS parameters negotiated with an external network.

The LTE QoS concepts inherits from the UMTS QoS concept described in [15] and presents many similarities in the principle. However, as the trend in LTE goes toward reducing the number of nodes and simplifying the networks as much as possible, the LTE QoS concept presents substantial changes compared to previous version [11].

The 3GPP QoS concept is based on the bearer principle. A bearer is a logical connection between two nodes insuring a certain service level characterized by a set of QoS parameters. The LTE bearer architecture is described in figure 1.3 [13].

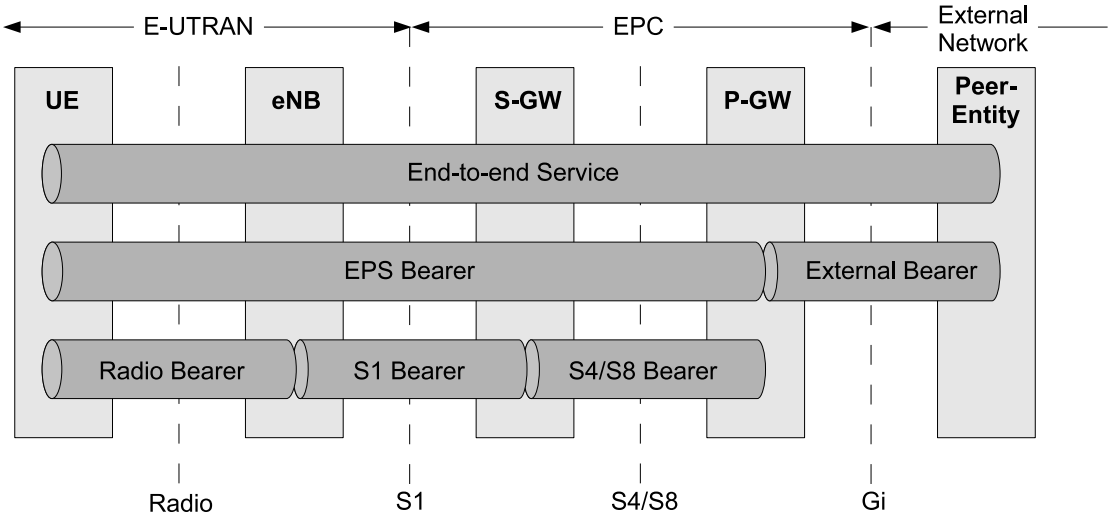

Figure 1.3: LTE bearer architecture. Reproduced from [13]

The main LTE QoS parameters described in [11] are the following.

- the Quality of Service Class Identifier (QCI) is a scalar that maps to a set of characteristics describing the expected packet forwarding treatment[16]. The QCI characteristics are detailed in table 1.1. New Channel Quality Information (CQI) can be configured by the operator. However 3GPP established a set standardized QCI that can be used as reference. Those QCI are described in table 1.2.

Table 1.1: QCI characteristics definitions [16]

Characteristic	Description
Resource Type	either "Guaranteed Bit Rate (GBR) bearer" or a "non-GBR bearer". GBR bearers provide the required GBR while non-GBR bearers don't provide any specific guarantee in terms of bit rate (best effor traffic)
Priority	1 corresponds to the highest priority. This parameter is to be used to differenciate bearers in case of resource shortage.
PDB	Packet Delay Budget. "Soft" upper bound with a confidence level of 98% for a time that a packet may be delayed between the P-GW and the UE.
PELR	Packet Error Loss Rate. Upper bound for the packet error loss rate of L2 SDUs.

Table 1.2: Standardized QCIs [16]

QCI	Resource Type	Priority	PDB	PELR	traffic type example
1	GBR	2	100ms	10^{-2}	Conversational Voice
2		4	150ms	10^{-3}	Conversational Video (Live Streaming)
3		5	300ms	10^{-4}	Non-Conversational Video (Buffered Streaming)
4		3	50ms	10^{-3}	Real Time Gaming
5	non-GBR	1	100ms	10^{-4}	IMS Signaling
6		7	100ms	10^{-3}	Voice, Video (Live Streaming), Interactive Gaming
7		6	300ms	10^{-4}	Video (Buffered Streaming), Transmission Control Protocol (TCP)-based
8		8			
9		9			

- the Allocation and Retention Priority (ARP) is used to prioritize bearers at bearer establishment (it may therefore be used in RAC),

Additionally, every GBR bearer is assigned the following parameters:

- the GBR,
- the Agregate Maximum Bit Rate (AMBR) which is the sum of the MBRs of all the bearers of a UE.

1.2.4 The Radio Protocol

Figure 1.4 presents the protocol stack of the E-UTRAN air interface between the eNode-B and the UE. At layer 3 (radio network layer) the Radio Resource Control (RRC) controls the establishment, performs maintenance and releases radio bearers. It also hosts mobility functionalities like handover and cell selection. In LTE, the RAC operates on radio bearer and is therefore at layer 3. Layer 2 (radio link layer) is divided into 3 sub-layers. The Packet Data Convergence Protocol (PDCP) performs Internetworks Protocol (IP) traffic specific tasks like header compression or duplicate detection. The Radio Link Control (RLC) hosts the Automatic Repeat ReQuest (ARQ) functionality and performs segmentation and concatenation of packets. The Medium Access Control (MAC) performs error correction through Hybrid Automatic Repeat reQuest (HARQ), UE prioritization and transport block format selection thanks to the LA functionality. The PS is located at layer 2. At layer 1 (radio physical layer) the PHYsical Layer (PHY) performs coding, Cyclic Redundancy Check (CRC) introduction and modulation. It hosts the LA functions and performs the CQI and HARQ Acknowledgement (ACK)/Non-Acknowledgement (NACK) reporting.

The LTE radio protocol has a channel structure. the different bearers are mapped into different logical channels defined by the type of data that are carried. The MAC performs mapping between logical channels and transport channels which are defined according to how the data are carried through the air interface. Finally, the PHY maps the transport channels into physical channels.

We note here that unlike in UTRAN, both RAC and PS are located in the same node. While in HSDPA, only the PS could benefit from the CQI reports, in E-UTRAN both the PS and the RAC can benefit from it.

1.3 Orthogonal Frequency Division Multiple Access

OFDMA is a multiple users radio access technique based on OFDM. OFDM consists in dividing the transmission bandwidth into several orthogonal sub-carriers each of which carries a different data stream. A set of modulated data symbols sent over the different sub-carriers consists of an OFDM symbol. In the present context, orthogonality between carriers means that the energy spectral density of the different sub carriers overlap without interfering with each other [17]. However, in a wireless channel, the signal can go through a multipath channel which may alter the subcarrier orthogonality at the receiver. For this reason, a cyclic prefix can be introduced before each OFDM symbol in order to conserve the orthogonality at the receiver [18] and therefore provide robustness in a wireless environment. Another key development of OFDM was the demonstration that the

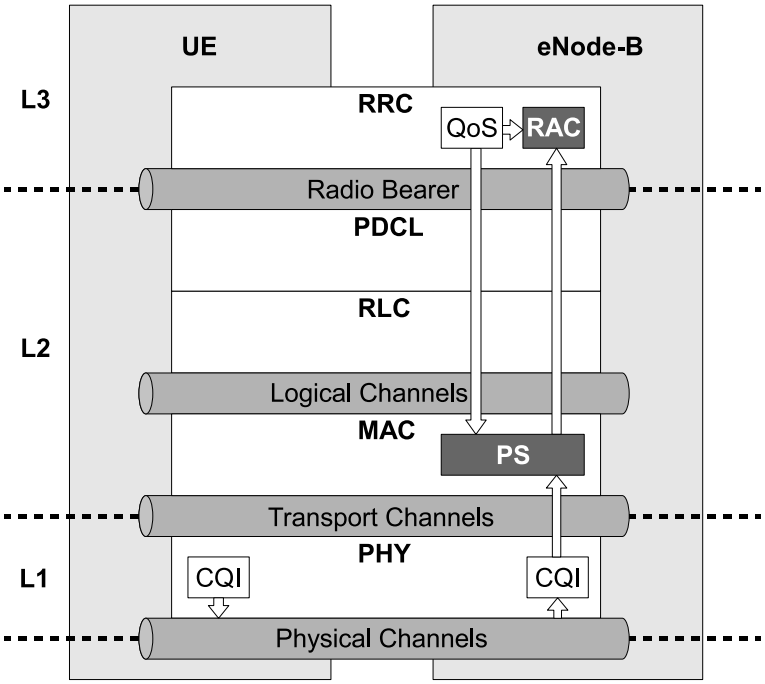

Figure 1.4: The radio air interface protocol stack

subcarrier modulation and demodulation can be performed by low complex algorithms IFFT and FFT [19]. Though invented in the 60s [17] OFDM became popular in the 90s [20]. Nowadays, OFDMA is often considered as the best choice for future broadband wireless access [6].

In a downlink context, OFDMA consists in sending data to different UEs using different sets of sub-carriers. In order to simplify and minimize the signaling of the resource allocation to the UEs, the sub-carriers can be grouped into different sub-sets each sub-set forming the smallest frequency domain allocation unit. In E-UTRAN, this unit is called Physical Resource Block (PRB) where each PRB consists of an equal number of adjacent sub-carriers. Similarly, OFDM symbols can be grouped by adjacent symbols to form the smallest time domain allocation unit which is call the Transmission Time Interval (TTI) in E-UTRAN. The domain division as well as the unit names employed in the thesis are summarized in figure 1.5. Figure 1.6 gives an example of 2 UE resource allocation.

In E-UTRAN, the sub-carrier spacing is 15kHz and each PRB consists of 12 adjacent sub-carriers. A TTI is composed of 14 OFDM symbols for the short cyclic prefix configuration and lasts for 1ms. More details regarding the OFDM / OFDMA parameters can be found in [21]. Moreover, LTE can operate on variable bandwidth as described in [22], table 1.3 summarizes the different configurations available.

1.4 Radio Resource Management: the State of the Art

As mentioned before, RRM is the set of components in the RAN that help achieving QoS while using efficiently the available transmission resource. The main RRM algorithms dealt with in the

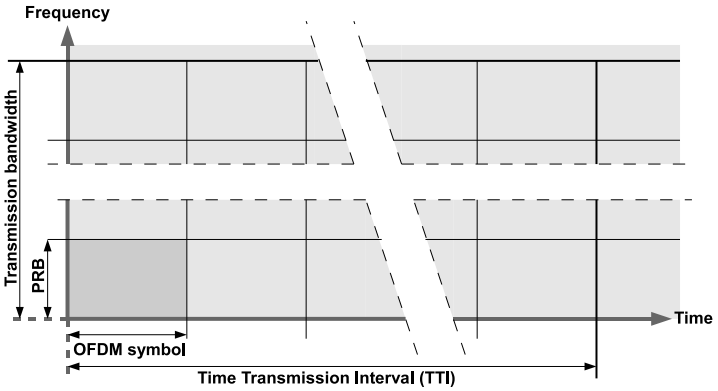

Figure 1.5: OFDMA transport units

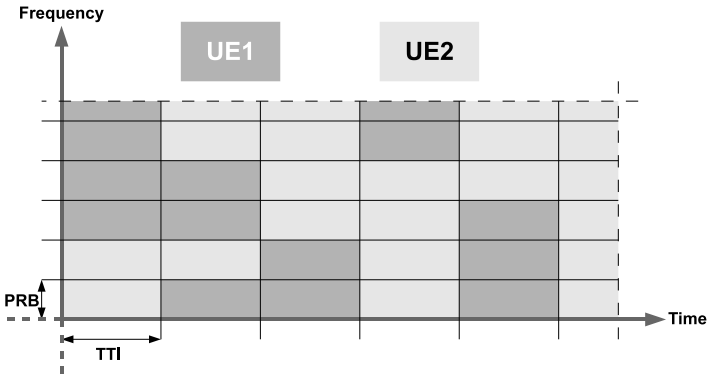

Figure 1.6: example of resource allocation in OFDMA

Table 1.3: E-UTRAN OFDMA configurations

Transmission Bandwidth (MHz)	1.4	3	5	10	15	20
Effective Bandwidth (MHz)	1.08	2.70	4.50	9	13.5	18
Number of PRBs (N_{PRB})	6	15	25	50	75	100

thesis are:

- The PS allocates radio resource to different services over a shared channel depending on the radio channel condition and the QoS parameters requirements,
- RAC aims at controlling the access of new services, insuring that the RAN is always able to provide QoS for all services,
- Inter-Cell Interference Coordination (ICIC) tries to insure that the generated inter-cell interference is as low as possible thus providing high channel signal quality.

1.4.1 Packet Scheduler

In a network node, the Packet Scheduler is the functionality in charge of prioritizing and forwarding data packets to the next nodes. The PS between a RAN and UEs over the air-interface takes a special importance due to the fast changing nature of the channel and the heterogeneity of the channel quality among UEs. In the most recent RAN like HSDPA or E-UTRAN, PS is located directly in the BS and is performed on a millisecond basis in order to adapt to fast channel variation and therefore benefit from multi user diversity gain which was first shown in [23].

Different optimization frameworks have been developed for wire-line networks and later generalized or adapted to wireless networks. For example, the service level agreement concept is based on the definition of service dependent network charging and incoming functions [24]. In that framework, the PS prioritizes UEs that will maximize the network income. Service level agreement has been adapted to OFDMA based wireless networks in [25]. Another example is the weighted fair queuing [26], which is derived from the general processor sharing concept and guarantees to different services a fixed share of the link capacity. Weighted fair queuing orders packets according to start and finish tags given to the packet at their arrival in the node. It was adapted to wireless networks in [27], [28], [29], [30] and further adapted to OFDMA wireless networks in [31].

Finally, the Proportional Fair (PF) packet scheduling algorithm proposed in [32] for wire-line networks is based on network utility maximization. Network utility maximization consists in defining the utility characterized by a function of the service data rate and then at each TTI the scheduling decision that maximizes the total network utility is taken. The expression of the decision can be found using a simple mathematical tool: convex optimization. In a wireless Time Division Multiple Access (TDMA) system, this results in a very simple expression where in each TTI, the UE that maximizes a simple metric is scheduled [33].

Following the utility maximization framework, many algorithms have been developed for TDMA wireless systems like HSDPA. The most significant example is the widely studied PF algorithms [34]. PF when in equilibrium state, schedules the UE in its highest fade. An interesting property of PF is that it provides an approximately equivalent share of the resource time-wise to each UE [35] [36]. As PF does not include any throughput guarantee, several improvements have been brought to the algorithms in order to add QoS support. Some examples are the PF with Required Activity Detection (RAD) algorithms and the PF with barrier functions algorithms.

PF with barrier function is introduced in [33] and [37]. It consists in introducing in the utility definition a "barrier function" that gives a very low utility to data rates below the minimum expected data rate. PF with RAD estimates the required time share required by the UE to fulfill its bit

rate requirement and adjusts the time share given to each UE based on their specific requirements [38].

The utility maximization framework can be applied using utility function of the packet delay [39]. The most widely studied algorithm is Modified Largest Weighted Delay First (M-LWDF) [40] [41] [42]. M-LWDF aims at guaranteeing to each UE that a defined proportion of their packets are delivered within a certain time laps.

Finally, many other PS have been developed for TDMA wireless networks. Some example are given in [43].

However, for OFDMA wireless networks, the application of the utility maximization principle is not as straightforward as in TDMA systems. Indeed, as different UEs can be allocated to different PRBs within the same TTI, the solution to the complex optimization problem does not necessarily apply to PS as a simple metric maximization [44]. PF for OFDMA wireless systems has been developed in [45] and [46]. As those solutions are computationally highly complex and could not be applied in a real system, [45] and [46] provide simplified solutions that reduce significantly the complexity of PF for OFDMA.

Many studies propose a mixed OFDMA PS approach where UEs are scheduled based on per-PRB metric maximization [47], [10]. This can also be considered as a simplification to the utility maximization problem. Moreover, some studies propose a PS divided into time and frequency domain packet scheduler [47]. The time domain scheduler first selects a set of UE to be scheduled and then, the frequency domain scheduler performs the UE to PRB mapping out of the preselected set of UEs. This offers very simple and flexible framework with very low complexity and therefore applicable in real products.

1.4.2 Radio Admission Control

Admission control consists in accepting new connections going through the node or not. As in wire-line networks, the capacity is constant and shared among connections, it is often possible to determine in a closed form the feasibility region of a PS algorithm. The feasibility region is the set of possible combination of connections that can be supported by the node while providing the required QoS to all connections. In that case, when a request for a new connection occurs, the admission control simply checks whether with the new connection the node would be in the feasibility region associated with the PS algorithm. If the node would remain in the feasibility region then the new connection is accepted. An example of this methodology can be found for the Earliest Deadline First and Static Priority PS algorithms in [48].

In Wireless system, the derivation of admission control (which will be from this point called Radio Admission Control (RAC)) algorithms is not that straightforward since the capacity of the channel is different among UEs and also varies. It is therefore not necessarily possible to determine a fixed feasibility region. Furthermore it is possible to predict precisely the capacity of incoming users in wire-line networks but this is not necessarily possible in wireless networks since it depends tightly on the knowledge of the radio conditions of the users, the extent of which depends on the signaling of the system before the establishment of the connection. RAC is therefore also very system dependent. In wireless systems, RAC generally focus on determining average limits of the feasibility zones.

For example, the algorithm proposed in [49] is based on the PF with barrier function PS al-

gorithm [33]. It tries to estimate the state of the system if the new connection is accepted and evaluate the penalty that the system would endorse. The admission control decision is then based on a penalty threshold. Another RAC algorithm has been proposed for HSDPA in [50]. This algorithm bases the RAC decision on the power required for a new bearer to support their defined QoS service level. A new bearer is accepted if the required power is available. Power is a convenient metric to use in HSDPA since power is reported to the RNC which hosts the RAC functionality. Furthermore, in HSDPA a direct coupling between RAC and PS is more complex to introduce since the two functionalities are not located in the same node.

RAC algorithms dedicated to OFDMA wireless system have been proposed in the literature. An interesting example is given in [51] which propose an a dynamic cell throughput estimation based RAC. The cell throughput estimate is used as upper bound for the sum of the GBR in the cell. In [52], a RAC based on the UE queue state is proposed. [53] proposes a simple carrier to interference ratio threshold RAC algorithm.

1.4.3 Inter-Cell Interference Coordination

ICIC is a mechanism that reduces the inter cell interference by coordinating the frequency use between neighboring cells and therefore aims at increasing the cell edge user capacity.

As ICIC techniques, several frequency reuse patterns for WiMAX are studied in [54]. Generally, using a fixed frequency reuse pattern with reuse factor lower than 1 will improve the channel conditions of the different UEs. However, this will limit the available bandwidth. Various types of soft frequency reuse schemes have been proposed in 3GPP for E-UTRAN. Soft frequency reuse consists in allocating different transmission power to different spectrum regions depending on the physical location of the antenna of the eNode-B so that every cell can benefit from a low interference spectrum region. This is of course at the expense of more interference on the other parts of the bandwidth. Several proposals are given in [55] and [56]. The performance of this type of scheme is studied in [57].

1.5 Motivation and Objectives

The general goal of the study is to provide a set of QoS aware RRM functionalities for OFDMA downlink. The algorithms design will be made keeping in mind that they must be realistic (low complexity) and implementable in a real system. Therefore the LTE specifications are taken as a case study. The objectives are divided into three main sub-goals.

Firstly, the study aims at designing a low complex QoS aware PS. Contrarily to many studies using idealistic assumptions, the algorithm design will include the main constraints related to a real system like noisy frequency domain channel quality reporting and HARQ handling. In order to insure low complexity, the PS algorithms will be built on the principle of decoupled time and frequency domain packet scheduler [47]. As literature is not very extensive on that type of scheduler, the design constraints specific to time / frequency domain scheduler will be studied. The expected outcome is a set of algorithms able to manage the scheduling of best effort and real time traffic mix with a maximum UE satisfaction.

Secondly, as explained in Section 1.4.3 OFDMA allows the use of frequency domain trans-

mission power variations. Transmission power variations can be used as a mean to increase the coverage of a cell. Besides, in limited offered traffic conditions, part of the spectrum can remain unused, which consists of another form of transmission power variation and can cause inconsistencies in the CQI mechanism. In the thesis, frequency transmission power variations are studied within a limited frame where the power spectral density can be either Nil or equal to a constant value. Cases where there exist parts of the spectrum where the power spectral density is null are called fractional load cases. The thesis aims firstly at showing what are the effects of a fractional load situation on the CQI reporting scheme and on the general QoS. Moreover, it aims at showing if the cell coverage can be increased by forcing fractional load in a cell (ICIC). The expected outcome are fractional load management techniques that add no implementation complexity and therefore don't use any communication between eNode-Bs.

The third objective of the thesis is to design RAC algorithms for OFDMA downlink to complement the QoS packet scheduler and therefore to help the management of best effort and real time traffic mix with a maximum UE satisfaction. It is intended to introduce new concepts with the new possibilities offered by LTE and OFDMA like the relocation of the RAC functionality at the eNode-B.

1.6 Methodology

A cellular network can be considered as a complex system as it involves a large number of dynamics and interactions. For example, within one cell, a PS decision will affect the state of all UEs present in the cell which will affect further the next PS decision. Another example is that the transmission power variation within one cell will affect the channel quality in neighboring cells, which will influence their RRM decisions and transmission power pattern thus affecting back other cells. Therefore, the simulation approach is the most appropriate to evaluate RRM algorithms in a realistic cellular system context. Furthermore, the performance of radio system depends on a certain number of non-deterministic phenomenons. The main example is the channel state. We therefore choose to evaluate the algorithms developed in the thesis in a system level simulator with the following characteristics:

- a multicell layout as well as the different layer of the E-UTRAN air interface protocol stack relating directly to RRM are implemented,
- non deterministic phenomenons or phenomenons that do not intervene into the interactions of the system are implemented with models of the literature or developed during the course of the thesis,
- link performance is evaluated in a separate simulator which is not part of the present study and is mapped into the system level simulator with techniques developed in the literature.

A special care is brought to the statistical relevance and significance of the simulation results by using appropriate statistical tools. Moreover, as part of the thesis work, a lot of efforts have been brought on the quality of the simulator in terms of computation time in order to enable performing simulations with the greatest possible statistical significance. For example, the channel implementation of the simulator has been optimized in order to reduce the computation time to the minimum and therefore enable to run longer simulations. However such a simulator remains heavy and the available computation power was limited during the PhD study. Some of the study cases

would require exceptionally long computation time to provide accurate performance evaluations. As such a time was not available during the PhD study, the simulation results provided can only be taken as performance indicators rather than precise evaluations. Those cases are mentioned explicitly in the thesis whenever they occur.

Analytical and statistical methods have been used when considered useful to define optimality criterions of RRM algorithms. Moreover, when possible, results are assessed by mean of analytical analysis considering simplified assumptions.

1.7 Novelty and Contributions

The overall contribution of the PhD work is the creation of a RRM concept for downlink OFDMA. The different novelties and contributions of the thesis can be divided into several distinct parts brought all together in chapter 6.

Firstly, the thesis proposes a consequent work regarding PS. Work has been carried out in collaboration with colleagues from Aalborg university and Nokia Siemens Networks. Collaborative contributions include early studies on multi user diversity gain from frequency domain PS for OFDMA and compressed CQI schemes. The main personal contributions consists of:

- The establishment of two design principles for decoupled time / frequency domain PS. Firstly the Independence of the time and frequency domain PS and secondly, the throughput control role should be given to the Time Domain Packet Scheduling (TDPS) in case of large number of UEs and to the Frequency-Domain Packet Scheduling (FDPS) in case of low number of UEs.
- Two FDPS metrics: Carrier over Interference to Average (CoItA) and Proportional Fair scheduled (PFsch). They provide a cell throughput gain of 10% over Throughput To Average (TTA)
- The metric weighting principle allows to introduce a low complex throughput control mechanism in the FDPS. Several weighting schemes are proposed. The concept proves to increase the cell coverage in case of low number of UEs in the cell.
- The design of the RAD scheduler for OFDMA. The design encompasses two different parts. The RAD TDPS can be further improved by introducing delay awareness. The RAD FDPS is based on the metric weighting principle.
- A thorough study of the following QoS aware TDPS: Priority Set Scheduler (PSS) and M-LWDF in OFDMA with decoupled time / frequency domain scheduling.

The different contributions are detailed in chapters 3 and 5 have been published in the following conference articles:

- A. Pokhariyal, K.I. Pedersen, G. Monghal, I.Z. Kovacs, C. Rosa, T.E. Kolding and P.E. Mogensen, "HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness", *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, pp. 2761-2765, Dublin, Ireland, April 2007.

- K.I. Pedersen, G. Monghal, I.Z. Kovacs, T.E. Kolding, A. Pokhariyal, F. Frederiksen and P.E. Mogensen, "Frequency Domain Scheduling for OFDMA withLimited and Noisy Channel Feedback", *Proceedings of the IEEE Vehicular Technology Conference (VTC)* Baltimore, USA, October 2007.
- G. Monghal, K.I. Pedersen, I.Z. Kovacs and P.E. Mogensen, "QoS Oriented Time and Frequency Domain Packet Schedulers for The UTRAN Long Term Evolution", *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, Singapore, May 2008.

Then, fractional load related contributions have been firstly made through collaborative work with the simple evaluation of the cell performance in fractional load conditions. the main contribution of the thesis regarding fractional load are

- A study of the consequences a fractional load situation due to lack of offered traffic. It is shown that it can lead to a dramatic BLock Error Rate (BLER) increase.
- New autonomous frequency transmission allocation patterns that overcome the above-mentioned BLER increase and minimize the interference at a given load.
- It is as well shown that a simple Wideband Interference Reporting in the CQI scheme overcomes the BLER increase.

The different contributions are detailed in chapter 4 have been published in the following conference articles:

- A. Pokhariyal, G. Monghal, K.I. Pedersen, P.E. Mogensen, I.Z. Kovacs, C. Rosa and T.E. Kolding, "Frequency Domain Packet Scheduling under Fractional Load for the UTRAN LTE Downlink", *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, pp. 699-703, Dublin, Ireland, April 2007.
- S. Kumar, G. Monghal, J. Nin, I. Ordas, K.I. Pedersen and P.E. Mogensen, "Autonomous Inter Cell Interference Avoidance under Fractional Load for Downlink Long Term Evolution", *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, Barcelona, Spain, April 2009.
- G. Monghal, S. Kumar, K.I. Pedersen and P.E. Mogensen, "Integrated Fractional Load and Packet Scheduling for OFDMA Systems", *Proceedings of the IEEE International Conference on Communications (ICC)*, Dresden, June 2009.

And the following patent application has been filled:

- K.I. Pedersen, P.E. Mogensen, G. Monghal and A. Pokhariyal, "Frequency-Domain Packet Scheduling under Fractional Load", United States Provisional Patent Application, April 2007.

RAC has been studied extensively and the main personal contribution are:

- A new RAC concept based on the RAD principle that takes both the channel quality of the UEs in the cell and that of the incoming UE.

- A thorough analysis of RAD-based RAC versus other algorithms of the literature concluding that the RAD-based RAC outperforms the other algorithms.

These contributions is detailed in chapter 6.

Additionally, the channel implementation optimization work has been formalized into a general OFDMA simulator channel optimization framework, where the main personal contribution consists of:

- a method for optimizing the fast fading implementation of an OFDMA fast fading simulator based on the frequency and time domain correlation properties of the channel.

This contribution is published in the following conference article:

- G. Monghal, I.Z. Kovacs, A. Pokhariyal, K.I. Pedersen, C. Rosa and P.E. Mogensen, "Fast Fading Implementation Optimization in an OFDMA System Simulator", *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, pp. 1214-1218, Dublin, Ireland, April 2007.

A reprint of this article can be found in appendix .

Besides, other topics closely related to the PhD study have been studied in collaboration with colleagues from Aalborg University and Nokia Siemens Networks. This includes LTE cell capacity with different sectorized cells and the impact of reduced CQI schemes. Those studies have been published in:

- S. Kumar, I.Z. Kovacs, G. Monghal, K.I. Pedersen and P.E. Mogensen, "Performance Evaluation of 6-Sector-Site Deployment for Downlink UTRAN Long Term Evolution", *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, Calgary, Canada, September 2008.
- K.I. Pedersen, G. Monghal, I.Z. Kovacs, T.E. Kolding, A. Pokhariyal, F. Frederiksen and P.E. Mogensen, "Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback", *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, Baltimore, USA, September 2007.

Finally, major contributions have been made to the company proprietary simulator in terms of implementation in the frame of the PhD study. Of course, the channel implementation optimization is part of it. Furthermore, the following functionalities have been implemented: major RRM features including the PS framework and RAC framework, CQI reporting and advanced CQI reporting schemes, Signal-to-Interference-plus-Noise Ratio (SINR) and CQI calculation, HARQ combining, Outer Loop Link Adaptation (OLLA), support of multi-traffic, Poisson call arrival, support of multiple site simulations and wrap around, support of partial bandwidth utilization and ICIC schemes including Soft Frequency Reuse (SFR) and simple frequency avoidance schemes.

1.8 Thesis Outline

The structure of the thesis follows a comprehensive structure where new degrees of complexity are added in every chapter. A brief description of the chapter is provided below:

- Chapter 1: *Thesis Introduction* provides information regarding the general background of the PhD Study. The technological context is described and the motivation for the PhD study is formulated. A description of the state of the art regarding RRM is included.
- Chapter 2: *System Description* describes the assumptions and the system model implemented in a simulator to test the different algorithms developed during the course of the study. The relevant assumptions and models are described with a high level of detail. Different traffic models and simulation scenarios are also introduced. Furthermore baseline results are provided in order to validate the simulator and to provide reference results.
- Chapter 3: *Packet Schedulers for throughput fairness and throughput control* introduces new PS with GBR guarantees. The underlying principles of the Packet Schedulers are discussed and we highlight the key mechanisms that help controlling the UE throughput and multi user diversity gain.
- Chapter 4: *Packet Scheduling Under Fractional Load Conditions* Fractional loading is defined as a situation where the transmission bandwidth is used only partially. The consequences of such a scenario are analyzed. Different scheduling algorithms dedicated to fractional load situations are developed according to different levels of channel signal quality awareness provided by the CQI. The scheduling algorithms are firstly tested in a simple scenario where the transmission bandwidth fraction in use is constant, then it is tested as a mean to increase coverage. Finally, is is tested in a realistic scenario where the fractional load situations are created by traffic fluctuation and lack of traffic.
- Chapter 5: *QoS aware Packet Scheduler* describes PS algorithms for QoS control. Algorithms from the literature and from Chapter 3 are adapted to support QoS. Furthermore, the RAD concept, well known to HSDPA systems, is adapted to OFDMA. All the algorithms are evaluated in mixed traffic conditions.
- Chapter 6: *Radio Admission Control* Introduces different RAC algorithms. The RAD concept is also adapted as a RAC algorithm. The different solutions are evaluated in a mixed traffic environment with advanced QoS aware PS.
- Chapter 7: *Thesis Conclusion* formulates recommendation for downlink OFDMA system RRM algorithms based on the PhD study. Furthermore, several points that could not be assessed during the thesis are discussed and proposed for future works.

In order to support the work, several appendix are provided:

- Appendix A: *Simulation Methodology: Simulator Optimization* includes a reprint of the article "Fast Fading Implementation Optimization in an OFDMA System Simulator", *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, pp. 1214-1218, Dublin, Ireland, April 2007.
- Appendix B: *Simulation Methodology: Statistical analysis of the results* presents the statistical framework used in the thesis to asses the statistical relevance of the different results.

Chapter 2

System Description

2.1 Introduction

Along the thesis, the different Radio Resource Management (RRM) algorithms are evaluated by means of system level simulations. In this chapter, we describe the general system model used during the thesis and implemented in the simulator. The different assumptions are detailed and their relevance to the study is discussed. The system model offers parameterization flexibility to study various RRM aspects. Firstly, we describe the physical characteristics of the system model, then we introduce the layer 3 and then layer 2 and 1 modeling of the system. Finally, baseline results are provided in order to provide reference to later results and to validate the simulator by comparing with results from the open literature.

2.2 Physical environment modeling

The system physical environment modeling is based on the MACRO case 1 assumptions defined by 3rd Generation Partnership Project (3GPP) in [9]. MACRO case 1 defines a geographical cellular configuration as well as pathgain, antenna pattern and a shadow fading model. It also includes several propositions for fast fading models, carrier frequencies and transmission configurations. All MACRO case 1 specific parameters and the fast fading models and transmission configurations chosen for the PhD study are detailed in Table 2.1 and in the rest of this section. The 3GPP model are used by different 3GPP participants to asses Evolved Universal Terrestrial Radio Access Network (E-UTRAN) performance and compare results. They are therefore widely used among researchers. This will help us validating the simulator by comparing the results with those given in other studies.

The simulated layout consists of 19 sites. Each site contains three sectorized antennas in its center oriented in different directions, each antenna is connected to a separate E-UTRAN Node B (eNode-B). A site consists therefore of three cells (alternatively called sectors) with independent RRM. The geometrical aspects of the layout are described in Figure 2.1.

The maximum transmit power per eNode-B, for a 10MHz transmission bandwidth E-UTRAN configuration, is of 46dBm. This 10MHz transmission bandwidth configuration comprises 600 sub-carriers. If all sub-carriers are in transmission mode with maximum and equal power, the

Table 2.1: MACRO case 1 system parameters

Parameter	**Setting**
Total number of sectors, N_{sec}	57
Sectors per site, $N_{secpsite}$	3
Site-to-site distance	500 m
Antenna pattern gain, $L^i_{ant}\left[\theta^i\right]$	$-\min\left\{12\cdot\left(\frac{\theta^i}{70^o}\right)^2, 20\right\}$ dB
Path Gain, $L^i_{path}\left[s\right]$, s in km	$-121.1 - 37.6\cdot\log_{10}\left(s\right)$ dB
Shadow fading standard deviation	8 dB
Shadow fading correlation (same site)	1
Shadow fading correlation (between sites)	0.5
Maximum sector transmit power P_{max}	46 dBm
Transmit power per sub-carrier	18.22 dBm
Thermal noise spectral density W_{th}	-174 dBm/Hz
UE noise figure W_f	9 dB
Fast Fading	Typical Urban 20, rayleigh
LTE System bandwidth configuration	10 MHz
Transmission bandwidth Bdw_{trans}	9 MHz
Number of sub-carriers	600
Sub-carrier spacing Bdw_{sub}	15 kHz
Carrier frequency	2 GHz

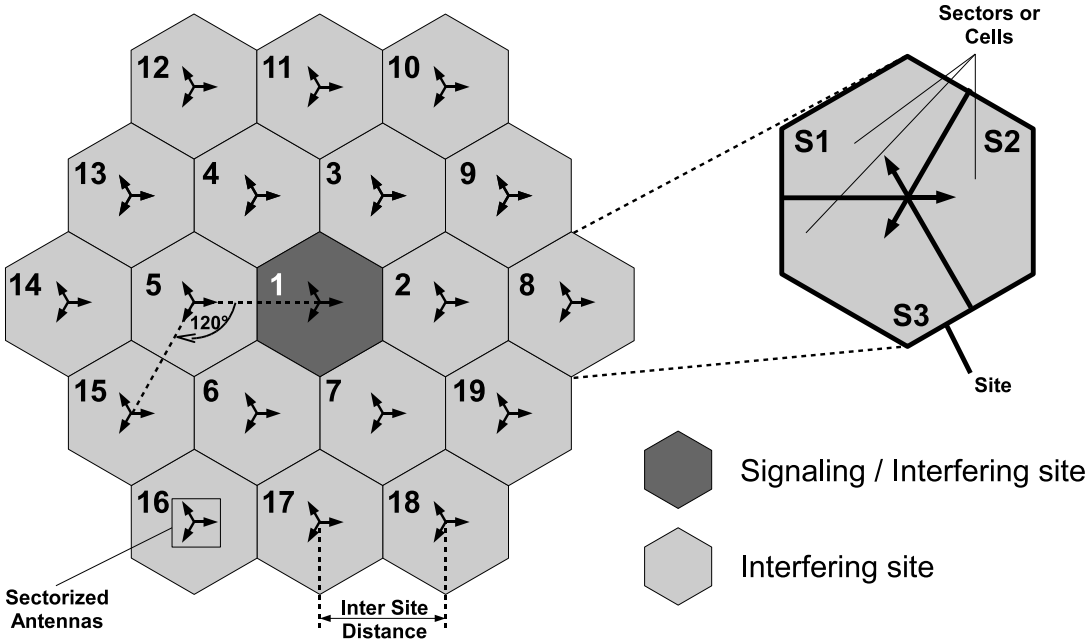

Figure 2.1: Cellular Layout Model

transmission power per sub-carrier is of 18.22dBm. In the PhD study, it is chosen that a sub-carrier can be:

- either in transmission mode, in which case, the sub-carrier is transmitting data with a power of 18.22 dBm
- or in non-transmission mode, in which case, the sub-carrier is not transmitting any data.

Along the thesis, two different layout simulation approaches depending on two different power transmission setting cases.

Firstly if all sub-carriers are in transmission mode at every Transmission Time Interval (TTI) in every cell, then the inter cell interference is constant. Therefore, as long as this condition is fulfilled, any RRM decision will have only internal impact on its own cell traffic as the inter cell interference remains unchanged. With that assumption, User Equipment (UE)s are dropped only in the three center cells called "signaling / interfering cells" where the RRM algorithms and data transmission are explicitly simulated. The other cells are distributed symmetrically around the center site and their purpose is only to create interference for the center site. Those cells are called "interfering cells". This layout simulation approach is illustrated in figure 2.1.

Secondly, if the transmission state of the sub-carriers is decided by the RRM algorithms, the transmit power of a cell can be subject to time and frequency variability. In that case, the inter cell interference will vary and therefore, RRM algorithms will have internal influence on the cell as well as influence on other cells under the form of interference variations. In that case, it is necessary to simulate explicitly RRM in all cells of the layout. However, in order to provide to each cell similar interference conditions, the wrap-around technique is used. Six mirror layouts are reproduced around the main 19 sites layout thus creating in total 7 versions of each site. Users are dropped only in the main layout however, the interference and signal strength are based on its virtual layout which consists of the closest version of each of the sites. This layout simulation approach is illustrated in Figure 2.2.

UEs are dropped in the simulated cells with uniform probability. When a UE is dropped, a link is created with each of the N_{sec} sectors. Each link with sector i is associated with a shadow fading S^i, antenna gain L^i_{ant} and a path gain L^i_{path} according to the models specified in Table 2.1. The UE establishes a signaling connection with the eNode-B S with highest overall path gain:

$$S = \arg\max_i \left(L^i_{ant} \cdot L^i_{path} \cdot S^i \right) \tag{2.1}$$

Each UE remains at the same location until its session ends. Therefore, the different path gains, antenna gains and shadow fading values remain constant until the end of the UE session. However, the UE is given a certain speed. Therefore fast fading values are changing accordingly. This type of simulations can be called "semi-static" as it takes into account the effect of movement on the fast fading variation but not on the shadow fading and path loss variations. The underlying assumption is that UEs are moving around the same approximate location.

The fast fading model employed in the simulations is the 20 path typical urban model [58] with Rayleigh fading. We also assume that multiple antennas on receivers (UEs) and transmitters (eNode-B) are uncorrelated and therefore, the different connections between different antennas of a transmitter-receiver couple fade independently. Fast fading values are generated with the Jakes fader described in [59]. In order to simulate Orthogonal Frequency Division Multiple Access

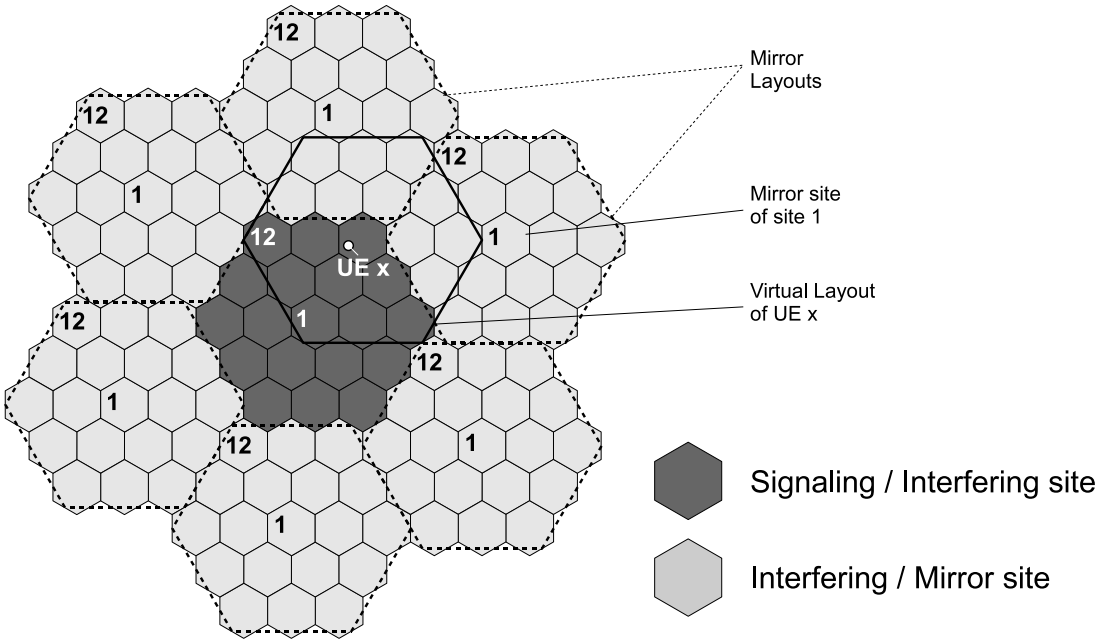

Figure 2.2: Cellular Layout Model

(OFDMA), fast fading gains $\mathbf{HH}^i\,[a]\,[k]$ have to be calculated for each sub-carrier k, each antenna connection a of each link i at every time unit which involves extensive use of Fourier transform. This can be very costly in terms of processing power. In order to limit the complexity of the simulator, the fast fading implementation has been optimized according to the method described in Appendix A. Furthermore, Appendix A describes in detail the calculation of $\mathbf{HH}^i\,[a]\,[k]$.

The system physical modeling allows to calculate different user's Signal-to-Interference-plus-Noise Ratio (SINR). In our simulations, transmissions use $N_{Tx} = 1$ transmit antenna and $N_{Rx} = 2$ receive antennas with Maximal Ratio Combining (MRC). The SINR provided by the MRC is approximated as follows in the simulations:

$$\mathbf{SINR}\,[k] = \sum_{a=1}^{N_{Rx}} \frac{\mathbf{P}^S\,[k] \cdot L^S_{ant} \cdot L^S_{path} \cdot S^S \cdot \mathbf{HH}^S\,[a]\,[k]}{\sum\limits_{i \neq S}^{N_{sec}} \left(\mathbf{P}^i\,[k] \cdot L^i_{ant} \cdot L^i_{path} \cdot S^i \cdot \mathbf{HH}^i\,[a]\,[k]\right) + W_f \cdot W_{th} \cdot Bdw_{sub}} \tag{2.2}$$

A frequently used general indicator that describes the average radio condition of a user is the Geometry (also called G-factor) defined as follows:

$$G = \sum_{a=1}^{N_{Rx}} \frac{P_{max} \cdot L^S_{ant} \cdot L^S_{path} \cdot S^S}{\sum\limits_{i \neq S}^{N_{sec}} \left(P_{max} \cdot L^i_{ant} \cdot L^i_{path} \cdot S^i\right) + W_f \cdot W_{th} \cdot Bdw_{trans}} \tag{2.3}$$

Table 2.2 summarizes the definitions of the different components used in the expressions of the SINR(2.2) and Geometry (2.3).

Table 2.2: Notations used in the expressions of SINR (2.2) and Geometry (2.3)

Variable	Unit	Description
$.^S$		superscript standing for the index of the signalink link
$.^i$		superscript standing for the index of interfering links
a		indicates the index of a receiver antenna
N_{Rx}		Number of receiver antennas (equal to 2 in the study)
N_{sec}		Number of sectors (equal to 57 in the simulated layout)
$\mathbf{P}^i[k]$	[W]	Transmission power of eNode-B i on sub-carrier k Sub-carrier transmission power is either -Inf. or 18.22dBm
$\mathbf{HH}^i[a][k]$	[-]	Fast fading gain on link i for receiver antenna a on sub-carrier k
L^i_{ant}	[-]	Antenna gain on link i
L^i_{path}	[-]	Path-loss on link i
S^i	[-]	Shadow fading on link i
W_f	[-]	Noise figure (?)
W_th	[W.Hz^{-1}]	Thermal noise (?)
Bdw_{sub}	[Hz]	subcarrier bandwith (15kHz)
Bdw_{trans}	[Hz]	effective transmission bandwidth (9MHz)

2.3 Call arrival and traffic models

Along the thesis we use four kinds of call arrival modes: infinite buffer, finite buffer, Poisson call arrival and constant User Diversity Order (UDO) with mixed traffic. Those different call arrival modes are used to provide different types of results. Those modes are thoroughly explained in the present section and we try to give some indication on their different purpose. Table 2.3 summarizes the different parameters to be set for each call arrival mode.

2.3.1 Infinite Buffer

Infinite buffer simulations consists of N_{run} runs of T_{IB} s. In each run, N_{UE} UEs per cell are dropped in the simulated layout. Each UE has an "infinite buffer" in the eNode-B to download. Therefore, a UE session ends with the end of the run. This type of simulations are used to understand basic features of Packet Scheduler (PS) algorithms. Infinite buffer simulations are easy to analyze and interpret since the time spend by each UE in the network is fixed and the number of UE is fixed as well.

Table 2.3: Parameters for traffic and call arrival settings

Configuration	Parameters
Infinite Buffer	Number of runs N_{run}
	Simulation Time T_{IB} [s]
	Number of UE per cell N_{UE}
Finite Buffer constant UDO	Simulation time T_{FB} [s]
	Number of UE per cell N_{UE}
	Buffer Size B_{FB} [kb]
Poisson Call Arrival	Simulation time T_{PCA}[s]
	Finite buffer average offered load $\overline{\gamma_{FB}}$ [kbps]
	Finite buffer Buffer Size B_{FB} [kb]
	Constant Bit Rate (CBR) average offered load $\overline{\gamma_{CBR}}$ [kbps]
	CBR packet size P_{CBR} [kb]
	CBR session time T_{CBR} [s]
	CBR bit rate CBR [bkps]
Constant UDO with mixed traffic	Simulation time T_{CUDO}[s]
	Number of finite buffer UE per cell N_{UE}^{FB}
	Finite buffer Buffer Size B_{FB} [kb]
	CBR offered load γ_{CBR} [kbps]
	CBR packet size P_{CBR} [kb]
	CBR session time T_{CBR} [s]
	CBR bit rate CBR [bkps]

2.3.2 Finite Buffer

Finite buffer simulations consist only of one run of T_{FB} s. In the beginning of the simulation, N_{UE} UEs per cell are dropped in the simulated layout. Each user has a buffer of B_{FB} kb to download. Once the buffer is fully downloaded, the UE is replaced by another UE dropped in the same cell so that the number of UE in the cell N_{UE} remains constant. The finite buffer call arrival mode is more fair and realistic than infinite buffer as each UE downloads the same amount of data.

2.3.3 Poisson Call Arrival

Poisson call arrival simulations consist in creating new UEs in the layout with a Poisson distributed time spacing between each new UE creation. In the Poisson call arrival mode, two different traffic types can be used. Firstly, a UE with finite buffer traffic model must download a buffer of size B_{FB} kb directly available at the eNode-B at UE dropping. The finite buffer model is used to

model users requiring best effort service in general like File Transfer Protocol (FTP) users of web browsing users. The other available traffic type is CBR. For a CBR UE, packets of equal size arrive at the eNode-B with a constant inter arrival time during a certain session time T_{CBR}. A CBR UE is entirely characterized by its bit rate CBR, session time T_{CBR} and packet size P_{CBR}. CBR traffic is used to model mainly video or music streaming users. In order to characterize the arrival rate of UEs in the layout, we use the Average Offered Load (AOL). This measure corresponds to the average data rate delivered to the base station in kbps. Poisson Call Arrival is the most realistic arrival scenario. Furthermore, it enables to test Radio Admission Control (RAC), which can be performed on any incoming UE.

2.3.3.1 Constant UDO with mix of traffic

Constant UDO simulations consists of one run of T_{CUDO} s. N_{UE}^{FB} finite buffer UEs and N_{UE}^{CBR} CBR UEs are dropped at simulation start. The number of finite buffer and CBR UEs remain constant until the end of the simulation. Therefore, every time a UE finishes its session, it is replaced by a UE of the same traffic type. Finite buffer UE can be characterized by N_{UE}^{FB} while CBR UEs can be characterized by the CBR offered load:

$$\gamma_{CBR} = N_{UE}^{CBR} \cdot CBR \tag{2.4}$$

2.4 The Radio Admission Control framework

In E-UTRAN, RAC applies on radio bearers. One UE can potentially transmit on several bearers at the same time. However, in our study, we limit to the case where UEs transmit only on one bearer. Therefore, from this point, we will only use the term UE. Figure 2.3 describes the general RAC model followed in the study. When a UE is incoming in the system and request a connection establishment, the layer 3 RAC functionality decides whether the connection is granted or if the UE is rejected.

The RAC functionality must evaluate if the system has enough resource available to support the incoming UE. For that purpose, Information about the state of the system is available. This encompass the different statistics or information that can be generated by the PS functionality and the Quality of Service (QoS) parameters of the UEs already present in the network with an active connection. Moreover, the RAC can use the incoming UE QoS parameters as well a channel quality indicator called Reference Signal Received Quality (RSRQ) that is transmitted on the Physical Uplink Control CHannel (PUCCH). In E-UTRAN, the RSRQ is the ratio between the received reference signal and the received signal on a certain section of the bandwidth [60]. We discuss models for the RSRQ estimation based on the G-factor in a later stage.

2.5 The Packet Scheduler Framework

2.5.1 Overview of the PS functionality and the data flow

Figure 2.4 gives an overview of the PS functionality relations with data flow and diverse control mechanisms used in our simulations.

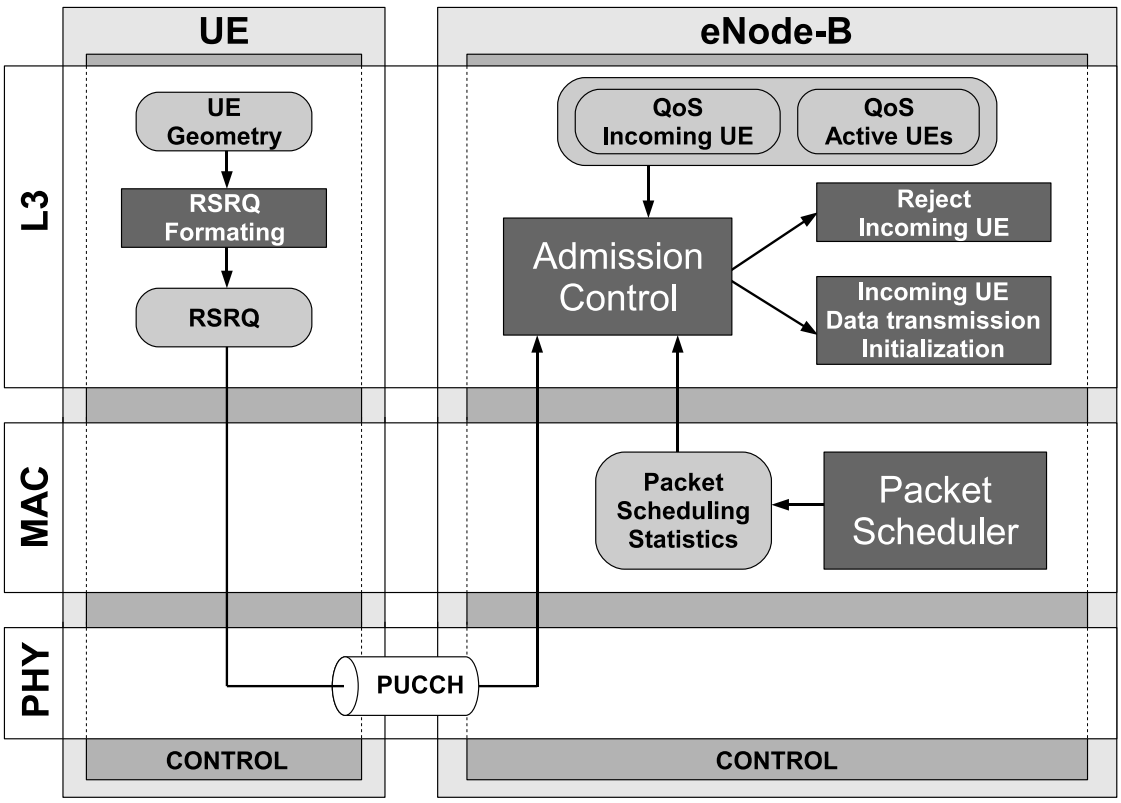

Figure 2.3: System Model, overview of Layer 3

Firstly, the data flow can be decomposed as follows. The eNode-B receives packets desalinated to each UE, which constitute a packet buffer for each UE. The Radio Link Control (RLC) performs segmentation and concatenation of those packets to create the Medium Access Control (MAC) transport block. Transport blocks are associated with a Hybrid Automatic Repeat reQuest (HARQ) channel that handles the retransmissions of the transport block in case of transmission failure. Transport blocks of each UE are given a certain Modulation and Coding Scheme (MCS) and Physical Resource Block (PRB) mapping and transmitted to each UE over the Physical Downlink Shared CHannel (PDSCH). Each UE decodes his received transport block. HARQ at the UE requests retransmission in case of transport block transmission failure. If the HARQ transmission succeeds, the RLC performs then de-segmentation and de-concatenation to deliver the original packets to the UE's upper layers. Note that we use a header overhead of 32 bits for the constitution of the RLC Protocol Data Unit (PDU) and 24 bits for the MAC PDU. According to the E-UTRAN specifications [61] [62], RLC and MAC header have a variable size depending on the size of the Service Data Unit (SDU), however, the chosen values correspond to an average value of the simulated cases. Furthermore, the functionalities of the Packet Data Convergence Protocol (PDCP) and the Automatic Repeat ReQuest (ARQ) functionality of RLC are not simulated as they have a low impact on the study.

In order to manage this data flow, several control mechanisms need to be in place. The core control mechanism is the PS which outputs the following informations:

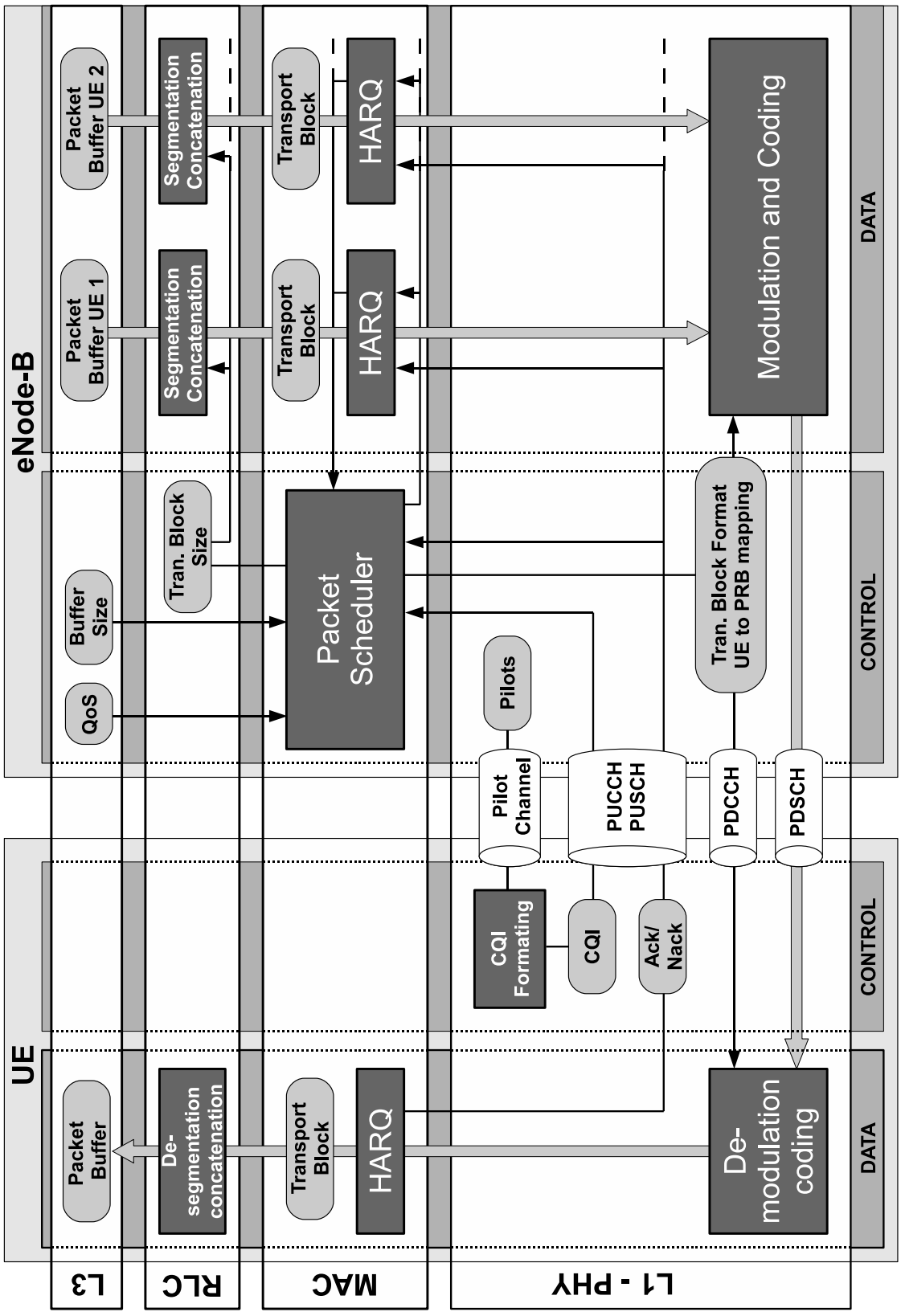

Figure 2.4: System Model, and overview of Layers 1 and 2

- *PRB to UE mapping* is the primary functionality of the PS. The following output are direct consequences of the PRB to UE mapping.
- *Transport block size* of the scheduled UEs are used by the RLC to perform concatenation and segmentation of the different packets in the UE packet buffer and create the transport block.
- *Transport block formats* of the scheduled UEs used by the PHYsical Layer (PHY) to perform modulation and coding.
- *scheduled HARQ pending transmission transport blocks* are passed by the MAC to the PHY together with the transport blocks for new transmission.

For providing these different informations, the PS can use the following different inputs:

- *The Channel Quality Information (CQI)* is reported on the PUCCH or the Physical Uplink Shared CHannel (PUSCH) by each UE on a periodic basis. It contains frequency domain channel quality information calculated based on the pilots.
- *The QoS parameters of the different UEs* can naturally be taken into account by the PS in order to provide he required service level to each UE.
- *The HARQ status of the UEs* is characterized by whether they have a pending retransmission or not. The PS can apply different degrees of prioritizing based on retransmission status.
- *The HARQ ACK/NACK reports* are reported on the PUCCH by the UEs every time a MAC transport block is received.
- *The buffer status of the different UEs* characterizes the total data that can potentially be transmitted per UE. It is also an upper bound for the transport block size.

2.5.2 Downlink physical channels

As illustrated on figure 2.4, several physical channels are considered in our study. We will describe in the present section the different modeling used for those different channels.

The PUCCH is not explicitely modeled in our study as we focus on downlink. We firstly assume that the PUCCH is error free, which is reasonable as the PUCCH is designed in the standard to be robust [21]. Secondly, we assume that the PUCCH has a sufficient capacity to accommodate all CQI and Acknowledgement (ACK) / Non-Acknowledgement (NACK) transmissions. The validity of this assumption depends directly on the CQI reporting scheme design discussed later and on the number of UEs present in a cell. A detailed study of the limitations related to uplink signaling channels is not the main point of the PhD study. However, as this is still a critical aspect that could be a bottleneck for the whole system, therefore collaborative studies on techniques aiming at reducing uplink signaling overhead have been led and are summarized in Appendix B.

Downlink channels are modeled with more details. Figure 2.5 illustrates the spacial configuration of the pilot channel, PDSCH and Packet Downlink Control CHannel (PDCCH) within a TTI.

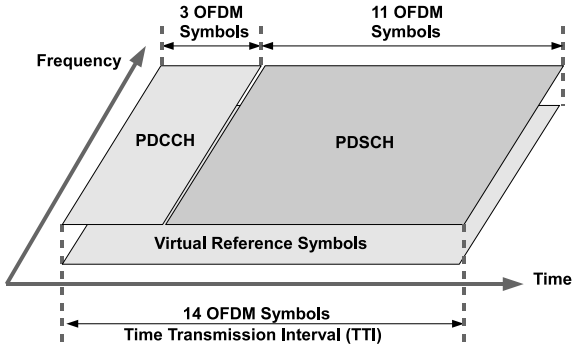

Figure 2.5: OFDMA transport units

In the PhD study, the PDSCH is using 11 of the 14 Orthogonal Frequency Division Multiplexing (OFDM) symbols contained in a TTI. When a UE is scheduled, it is transmitting a transport block over a subset of PRBs with a certain MCS. At each TTI, a block error probability is calculated for each scheduled UE following the Exponential Effective SINR Metric (EESM) method detailed in [63]. It consists in calculating an effective SINR:

$$\mathbf{SINR_{eff}}\left[n\right] = \beta_{MCS(n)} \exp\left(\frac{1}{\#\text{Sub}\left(n\right)} \sum_{k \in \text{Sub}(n)} \exp - \frac{\mathbf{SINR}\left[k\right]}{\beta_{MCS(n)}}\right) \tag{2.5}$$

In Formula (2.5), the effective SINR of UE n is calculated depending on:

- a β factor that depends on the MCS used by UE n: $MCS\left(n\right)$.
- the different SINR values of the different sub-carriers on which UE n is scheduled. In the formula, the set of sub-carriers used by UE n is denoted by Sub $\left(n\right)$.

The effective SINR is then mapped to block error probability by look-up into MCS dependent SINR to bloc error probability tables also called Actual Value Interface (AVI) tables. The reception of the transport block is then decided with the help of the block error probability and a uniformly distributed random variable generator. The application of the EESM method requires prior extensive link studies in order to generate:

- The MCS dependent β factors necessary for calculating the effective SINR
- The AVI tables

The link studies are not part of the PhD study. Tables generated by Nokia Siemens Networks (NSN) colleges have been used.

The pilot channel consists of virtual pilots spread over 14 OFDM symbols and over the whole bandwidth. Pilots are virtual as they are superposed with PDSCH data symbols. Our modeling of the pilot channel consists therefore of an idealization since the pilot overhead is not taken into account and a realistic pilot configuration would not cover the whole bandwidth. The pilot power per sub-carrier is $P^P = 17dBm$ which corresponds to 44.77dBm for the whole bandwidth.

The PDCCH is located on the 3 first OFDM symbols of the TTI in accordance to Long Term Evolution (LTE) specifications [21]. It is not explicitely modeled as the pilot channel or the PDSCH. As the PUCCH, the PDCCH is considered to be error free and non capacity limited. The validity of the second assumption depends on the maximum number of UEs scheduled during the same TTI. This consideration is taken into account in the PS design. More can be found on the limitations related to PDCCH in [64].

2.5.3 Hybrid Automatic Repeat reQuest

HARQ is the MAC layer retransmission procedure. In E-UTRAN downlink, HARQ is asynchronous; which means that retransmission can occur at any time. Furthermore, HARQ is adaptive: retransmissions don't necessary use the same PRBs nor the same MCS. However, In our study, the same MCS is always used in retransmissions of the same transport block. Using the same MCS helps reducing the signaling overhead and simplifies the HARQ soft combining implementation. It is therefore more likely to be implemented in a real product. Finally, in the study, a UE cannot transmit new data and a retransmission at the same time.

Each UE is given 6 stop-and-wait channel upon connection to the eNode-B. When a transport block is created by the MAC sub-layer, it is associated to an HARQ channel if any is available. Each channel performs transmission and retransmissions of the transport block upon command of the PS. A retransmission can be triggered only when a NACK has been received from previous (re)transmission. An ACK ends the HARQ process successfully. No more than 4 transmissions of the same transport block can be performed. If the transmission is not successful after 4 transmission, the transport-block is considered as lost. The ACK/NACK reporting delay is 2 TTI.

Chase combining is performed on the multiple received versions of the transport block. Our HARQ combining modeling is based on [65] with a chase combining efficiency of 1.0.

2.5.4 Channel Quality Information reporting

In order to provide frequency domain channel quality indication to the PS, a CQI is reported from each UE to the eNode-B on a periodical basis. In our study, the CQI is reported every 5 TTI with a delay of 2 TTI.

The CQI consists of a set of values corresponding to an estimate of the SINR on each CQI block. In our study the CQI block size is of 2 PRBs equals to $N_{CQI} = 24$ sub-carriers. The expression of the ideal CQI on m^{th} CQI block is as in (2.6):

$$\mathbf{CQI}^{id}[m] = \sum_{a=1}^{N_{Rx}} \frac{\sum_{k=(m-1)\cdot N_{CQI}+1}^{m\cdot N_{CQI}} \mathbf{P}^{S}[k][k] \cdot L^{S}_{ant} \cdot L^{S}_{path} \cdot S^{S} \cdot \mathbf{HH}^{S}[a][k]}{\sum_{k=(m-1)\cdot N_{CQI}+1}^{m\cdot N_{CQI}} \left(\left(\sum_{i\neq S}^{N_{sec}} \mathbf{P}^{i}[k] \cdot L^{i}_{ant} \cdot L^{i}_{path} \cdot S^{i} \cdot \mathbf{HH}^{i}[a][k] \right) + W_f \cdot W_{th} \cdot Bdw_{sub} \right)} \tag{2.6}$$

Note that the CQI is an SINR measure on each CQI block. The different notations of (2.6) are summarized in Table 2.2. Receiver imperfections are modeled by adding a zero mean Gaussian error of 1dB standard deviation to the ideal CQI as in [66]:

$$\mathbf{CQI}_{dB}^{er}[l] = \mathbf{CQI}_{dB}^{id}[l] + \epsilon \tag{2.7}$$

where ϵ is normally distributed with mean 0dB and standard deviation 1dB. In order to be implementable in a real system and in order to comply with the uplink signaling constraint mentioned in section 2.5.2, the CQI is further quantified with a 1dB step:

$$\mathbf{CQI}_{dB}^{trans}[l] = \lfloor \mathbf{CQI}_{dB}^{er}[l] \rfloor \tag{2.8}$$

Finally, the CQI needs time to be processed at the UE and to be transmitted back to the eNode-B. Therefore, it is delivered with a delay of two TTIs after measurement.

2.5.5 Packet Scheduler

Figure 2.6 describes in detail the PS framework used in the PhD study. The two main entities of the PS namely the Time Domain Packet Scheduling (TDPS) and the Frequency-Domain Packet Scheduling (FDPS) take as input a set of schedulable users. The schedulability is checked based on the availability of HARQ process and the availability of buffered data per each UE. In order to perform the scheduling decision, TDPS and FDPS can use the L3 QoS attributes of each UE. Furthermore the Link Adaptation (LA) and Outer Loop Link Adaptation (OLLA) functionalities can provide an estimate of the achievable throughput on any sub-band for each UE based on the latest CQI and ACK / NACK reports. Finally, TDPS and FDPS can use the HARQ status of the different UEs in order to apply different degrees of prioritizing to UEs with and without pending retransmission.

The TDPS firstly selects a subset of N_{mux} schedulable UEs and passes them to the FDPS which determines transport block size, PRB to UE mapping, MCS, and which HARQ process of each UE to transmit. the preselection of N_{mux} UEs has two important aims. Firstly, as the PDCCH has a limited capacity, we need to limit the number of UEs scheduled in order to comply with our assumption in section 2.5.2. According to [64], $N_{mux} = 10$ is a value that matches the PDCCH capacity. Furthermore, as the PS acts on a TTI basis, it is important to keep the PS to a low complexity in order to provide realistic computation time implementable in a real product.

2.5.6 Link Adaptation and Outer Loop Link Adaptation

LA provides an estimate of the achievable throughput on any sub-band and the MCS that must be used for that purpose. In order to provide that estimate, LA calculates the effective CQI based on the EESM method on the desired sub-band. By look-up in AVI tables, the LA finds the highest MCS that complies with the system BLock Error Rate (BLER) target and returns the corresponding achievable throughput.

The effective CQI is calculated based on the different CQI values adjusted by the OLLA offset:

$$\mathbf{CQI}_{dB}^{us}[l] = \mathbf{CQI}_{dB}^{trans}[l] - \Delta_{OLLA} \tag{2.9}$$

As can be seen in (2.9), only one offset is used per UE. The OLLA offset value is calculated according to the method described in [67]. The offset is initialized to Δ_{init} and is updated every time the UE receives an acknowledgment report of a first transmission of an HARQ process:

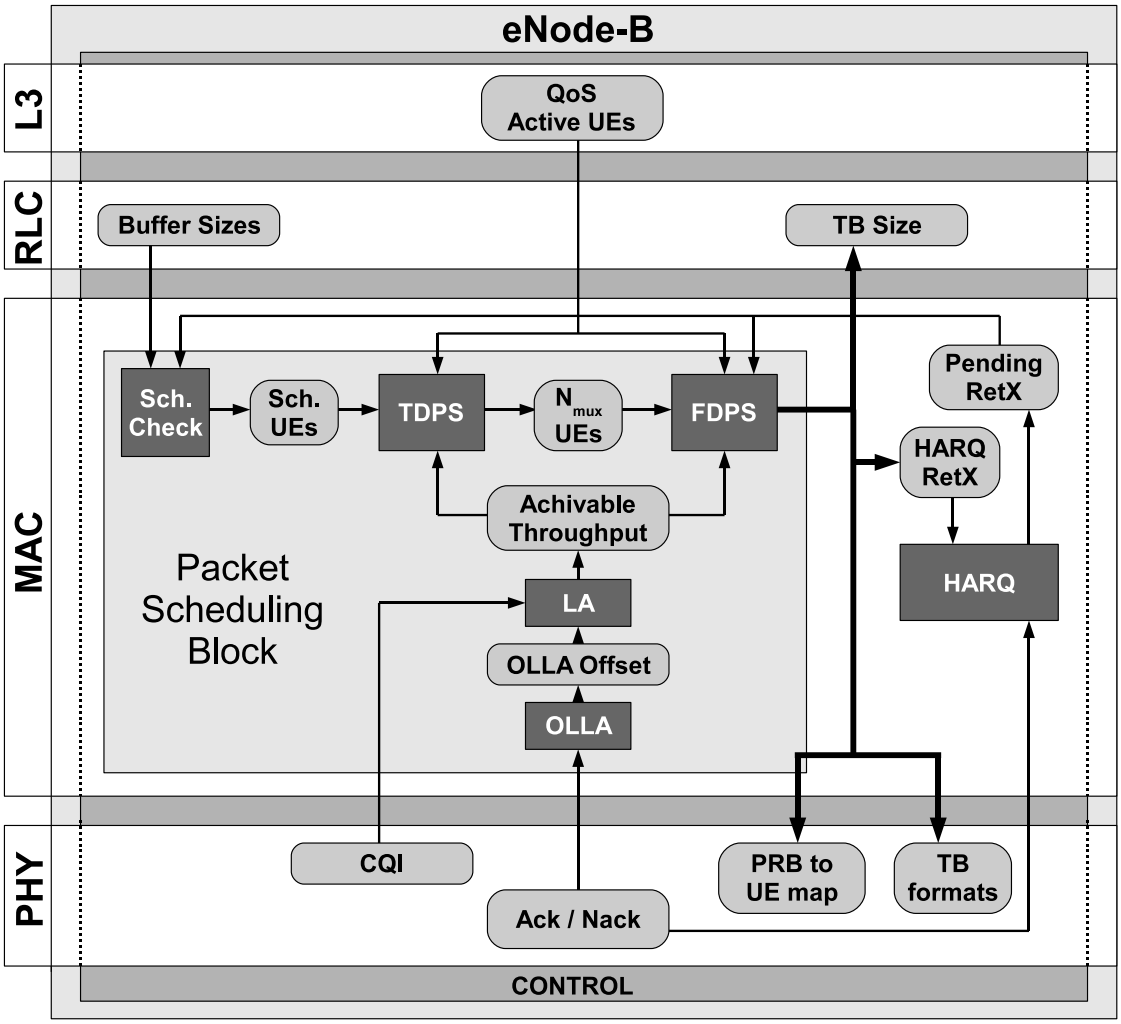

Figure 2.6: Packet Scheduling Model

- if ACK is received for first transmission, then Δ_{OLLA} is decreased by Δ_{down}:

$$\Delta_{OLLA} = \Delta_{OLLA} - \Delta_{down} \tag{2.10}$$

- if NACK is received for first transmission, then Δ_{OLLA} is increased by Δ_{up}:

$$\Delta_{OLLA} = \Delta_{OLLA} - \Delta_{up} \tag{2.11}$$

The relation between Δ_{down} and Δ_{up} is derived as follows:

$$\Delta_{down} = \Delta_{up} \cdot \frac{BLER}{(1 - BLER)} \tag{2.12}$$

2.6 Baseline results

In this section base line results are provided with the most common schedulers in the literature namely the Round Robin (RR) scheduler and the Proportional Fair (PF) scheduler. Those results can be used as reference for the rest of the thesis. Moreover, this will help us validate the simulator versus other results in the literature. We will give an analysis providing the basic understandings of the system and at the same time highlighting the different challenges. The main performance indicators used along the thesis will also be introduced and defined.

2.6.1 The Round Robin Scheduler

The RR scheduler consists in scheduling UEs in ordered turn with equal quantity of resource, disregarding channel quality or and priority criterion. We give here a more precise definition adapted to our general PS framework encompassing the split into time and frequency domain PS.

RR TDPS: Each UE is given a sequence number n. When a UE is removed, the sequence numbers are rearranged so that all UEs keep the same sequence order. When a UE is incoming in the cell, the sequence numbers are rearranged so that the new UE enters at the head of the sequence. Every TTI, the next N_{mux} UEs of the sequence that are scheduled and passed to the FDPS.

RR FDPS: A group of $\frac{N_{PRB}}{N_{mux}}$ adjacent PRBs are allocated to the UEs. The order of the allocation is made according to increasing sequence number. Note that $N_{mux}|N_{PRB}$ is a necessary condition for the scheduler definition to be valid. RR FDPS does not include any prioritizing considering the HARQ status of the UEs.

2.6.2 General HARQ process prioritizing

The RR schedulers does not perform any prioritizing depending on the HARQ status, however once the PRB to UE mapping is performed the different HARQ process of each UE are prioritized as follows:

1- Pending retransmissions: From the oldest HARQ process to the most recent HARQ process

2- New transmission: Create new HARQ process

This prioritizing strategy is used in the entire PhD study.

2.6.3 Metric based PS - the Proportional Fair scheduler

The version of PF we will introduce here is a low complex version used in other studies like [47] [68]. It is a generalization from the standard PF metric based scheduling for time domain multiple access systems like High-Speed Downlink Packet Access (HSDPA) [35]. It contrasts with the PF schedulers introduced in [45] and [46] where PF rigorously maximizes the logarithmic utility functions in an OFDMA context. In the following, we introduce the general metric based PS framework and define the different PF metrics. The metric based PS framework is later reused together with the definitions of new metrics.

Metric based TDPS: At a given TTI t, every UE n is associated with a metric $\mathbf{M}^{TD}[t][n]$. The N_{mux} UEs that maximize the metric are passed to the FDPS. Note that if the system has less than N_{mux}, then the TDPS is inactive.

PF TDPS: The time domain PF metric is defined as follows:

$$\mathbf{M}_{PF}^{TD}[t][n] = \frac{\hat{\mathbf{D}}[t][n]}{\overline{\mathbf{R}}[t][n]} \tag{2.13}$$

where $\hat{\mathbf{D}}[t][n]$ is the estimated wideband throughput of UE n at TTI t such as given by the LA functionality. $\overline{\mathbf{R}}[n,t]$ is the past average throughput defined by exponential averaging:

$$\overline{\mathbf{R}}[t][n] = \frac{T-1}{T} \cdot \overline{\mathbf{R}}[t-1][n] + \frac{1}{T} \cdot \mathbf{R}[t-1][n] \tag{2.14}$$

where $\mathbf{R}[t-1][n]$ is the scheduled throughput of UE n at TTI $t-1$. Note that if UE n is not scheduled at TTI $t-1$ then $\mathbf{R}[t-1][n] = 0$. T is the exponential filter constant. It is an important factor as it defines the steadiness and the speed of convergence of $\overline{\mathbf{R}}[t-1][n]$ if it converges.

Metric based FDPS: At a given TTI t, every UE, PRB couple p, l is associated with a metric $\mathbf{M}^{FD}[t][p,l]$. The PRB allocation takes into account the HARQ status of the N_{mux} selected UEs following the method described in [68]. The principle consists in allocating first the UEs without retransmission to the best PRBs and then allocating the remaining PRBs to the UEs with retransmissions. The reason is that retransmissions benefit from combining, and therefore don't need to be given the best PRBs. The general metric based algorithms of [68] has been improved to support cases with limited transmission buffer. Here follows a precise description of our metric based algorithm.

1 *Calculate the number of PRB x required for the various retransmissions.* To UE p, x_p is

- the number of PRBs associated with the oldest pending retransmission for UEs with pending retransmissions
- 0 for UE without pending retransmission

then

$$x = \sum_{p=1}^{N_{mux}} x_p \tag{2.15}$$

2 *Allocate UEs without retransmission.* The set of schedulable UEs without retransmission S_{sch}^{NoRetx} contains all the UEs without pending retransmissions among the N_{mux} UEs preselected by the TDPS. The set of schedulable PRBs S_{sch}^{PRB} consists of all the PRBs and the number of scheduled PRB N_{schPRB} is equal to 0.

- While S_{sch}^{NoRetx} is not empty and $N_{schPRB} < N_{PRB} - x$ do the following:
  - Schedule UE p' on PRB l' so that:
    - UE p' belong to S_{sch}^{NoRetx}
    - UE l' belong to S_{sch}^{PRB}
    - given those two conditions, $\mathbf{M}^{FD}[t][p', l']$ is the highest value of matrix $\mathbf{M}^{FD}[t]$.
  - Remove PRB l' from S_{sch}^{PRB}.
  - Calculate with LA the potential transport block size for UE p' based on the PRBs that are already allocated to UE p'.
  - If the transport block size is equal or greater than the data available in the buffer of UE p' then, remove UE p' from S_{sch}^{NoRetx}.

3 *Allocate UEs with retransmission.* The set of schedulable UEs with retransmission S_{sch}^{Retx} contains all the UEs with at least one pending retransmission among the N_{mux} UEs preselected by the TDPS.

- While S_{sch}^{Retx} is not empty do the following.
  - Schedule UE p' on PRB l' so that:
    - UE p' belongs to S_{sch}^{Retx}
    - UE l' belongs to S_{sch}^{PRB}
    - given those two conditions, $\mathbf{M}^{FD}[t][n', l']$ is the highest value of matrix $\mathbf{M}^{FD}[t]$.
  - Remove PRB l' from S_{sch}^{Retx}.
  - If UE p' has been allocated $x_{p'}$ PRBs then, remove UE n' from the set of schedulable UEs.

PF FDPS: The frequency domain PF metric is defined as follows:

$$\mathbf{M}_{PF}^{FD}[t][p,l] = \frac{\hat{\mathbf{d}}[t][p,l]}{\overline{\mathbf{R}}[t][n]} \tag{2.16}$$

where $\hat{\mathbf{d}}[t][p,l]$ is the estimated throughput of UE p on PRB l at TTI t such as given by the LA functionality.

2.6.4 Why a Decoupled Time / Frequency Domain Packet Scheduler?

The main reason behind introducing a decoupled time and frequency domain PS is the complexity reduction. Indeed, the TDPS preselects a subset of the UEs to be scheduled. This operation is made with a relatively low complexity since the time domain scheduler needs only calculate one metric per UE. On the contrary the FDPS needs to calculate N_{PRB} metrics per UE, therefore, the preselection made by the TDPS can greatly decreases the calculation complexity taken by the FDPS.

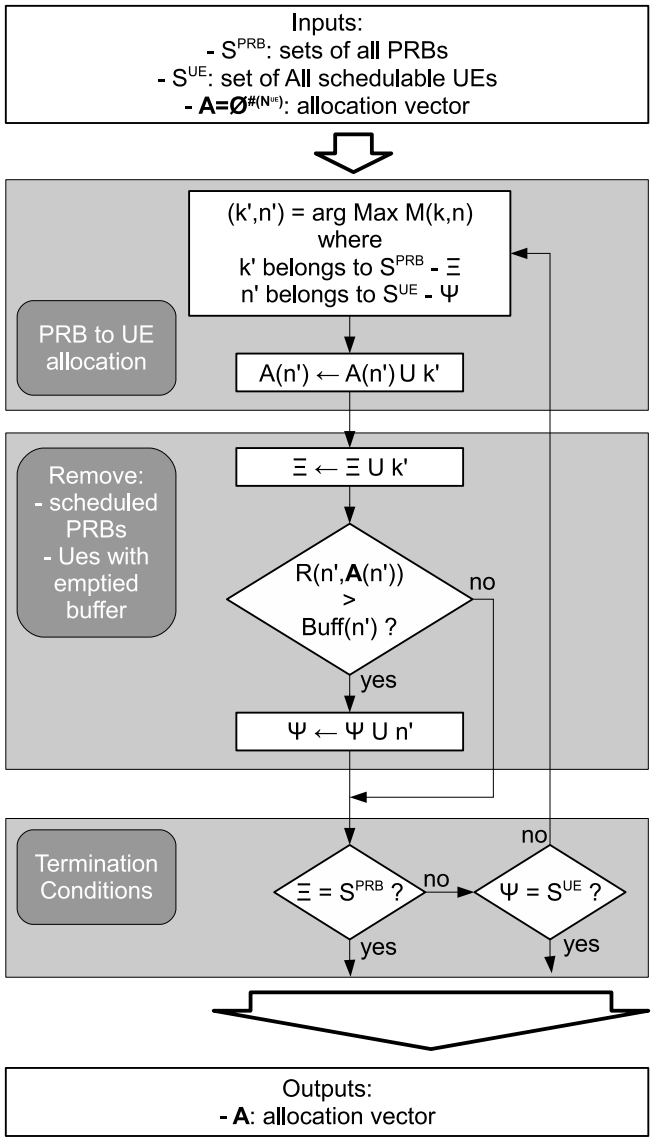

Figure 2.7: Description of the modified PRB allocation algorithms with integrated PRB prioritizing

The complexity (in number of calculated metrics) of a system with only FDPS would be equal to:

$$Comp_{FDPSOnly} = N_{UE} * N_{PRB} \tag{2.17}$$

while the complexity of a system with a decoupled TDPS / FDPS is equal to:

$$Comp_{TDPS/FDPS} = N_{UE} + (\min(N_{MUX}, N_{UE}) * N_{PRB}) \tag{2.18}$$

Figure 2.8 compares the two different complexity curves.

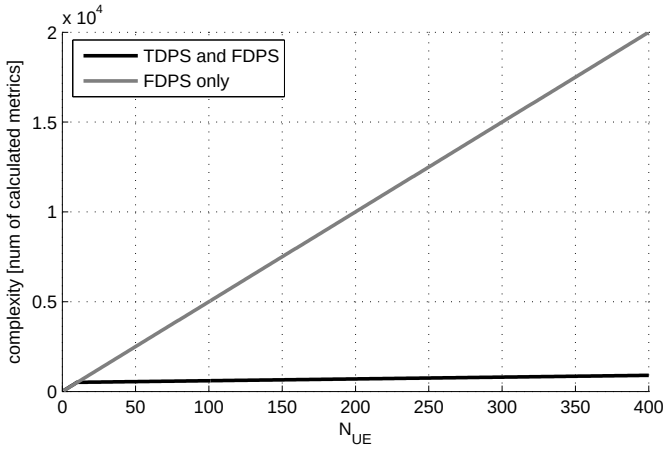

Figure 2.8: compared complexity of a PS with only FDPS and with a decouple TDPS / FDPS

2.6.5 Results

Table 2.4: Baseline simulation parameters

Parameter	**Setting**
Envirronment	MACRO #1
Layout Configuration	1 simulated site
Traffic models / call arrival	Infinite Buffer N_{run}=50 T_{IB}=10s N_{UE}=30
	Finite Buffer T_{FB}=500 N_{UE}=30 B_{FB}=3.33Mbits
Packet Schedulers	RR PF

The specific simulation parameters for the baseline results are summarized in Table 2.4. Simple results with finite buffer and infinite buffer traffic model allow us to analyze and understand

the basic functionalities of the system. As they are also the most widespread assumptions in early E-UTRAN studies, they will allow us to validate our simulator by comparing our results with other studies of the literature.

2.6.5.1 Key Performance Indicators

Along the thesis, we use similar key performance indicator to characterize the performance of the system under different assumptions and algorithms. We define the main key performance indicators in the present section. Later in the thesis, new key performance indicators a accompanied by a definition.

The UE throughput corresponds to the throughput generated by HARQ acknowledged bits at the PDCP layer over a whole UE session. This means that the overhead bits due to MAC PDU and RLC PDU headers are not taken into account. The UE throughput is often presented as a Cumulative Density Function (CDF).

The Cell coverage is defined by the UE throughput at 5% outage. This measure is widely used in the litterature and is an indicator of the throughput offered to cell edge UEs.

The cell throughput R_{cell} is the PDCP layer HARQ acknowledged throughput transmitted by an eNode-B over a simulation. As the UE throughput, the cell throughput does not take into account the overhead bits due to MAC PDU and RLC PDU headers. As several eNode-B are simulated together. The cell throughput is presented as average cell throughput $\overline{R}_{cell}$ which is the arithmetic average over all the simulated cells during one simulation.

2.6.5.2 Baseline results and comparison with other studies

Figures 2.9 and 2.11 show the average cell throughput and the cell coverage. In the infinite buffer case, the cell throughput is of 14.27 Mbps with the PF scheduler and the coverage is of 189 kbps. Several studies have been published involving similar assumptions. We will focus here on the studies published in 3GPP documentation [69] and [70]. [70] has been further published in a conference in [71].

With similar assumptions and similar simulation methodology, [69] shows a cell throughput of 14.92 Mbps (+4.8% compared to our results) while [70] shows 13.8 Mbps (-3.1% compared to our results). Those results show that our simulator is in line with external studies. the 5% difference can be explained by several reasons mentionned by order of importance:

- It is not mentioned in other studies what PDCCH overhead neither what MAC and RLC header is considered . In our study for example, the PDCCH overhead accounts for 21.43% of the bandwidth. Therefore, a difference in that value can have a significant impact on the results.
- The implementation of the MAC layer functionalities like PS, LA and HARQ is not described in details. As those functionalities are not specified in the 3GPP standard, there may be differences compared to our implementation described in section 2.5.
- The CQI reporting scheme is not mentioned either and can be a factor. For example, using ideal CQI reports containing exact SINR information on each PRB could show better results.

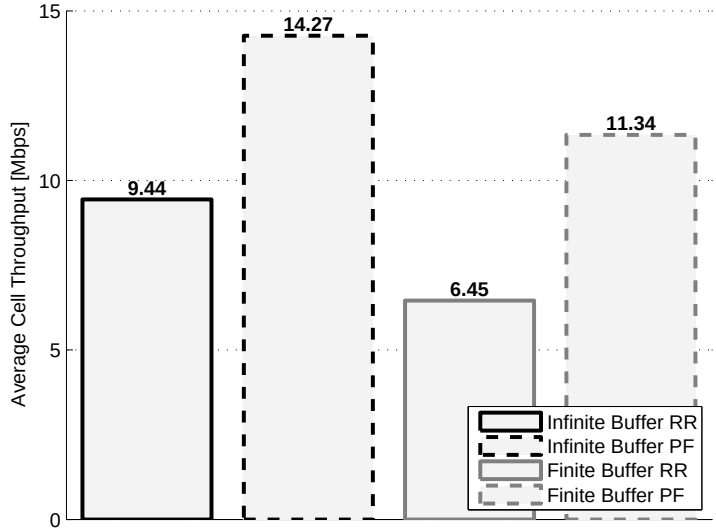

Figure 2.9: Average Cell Throughput

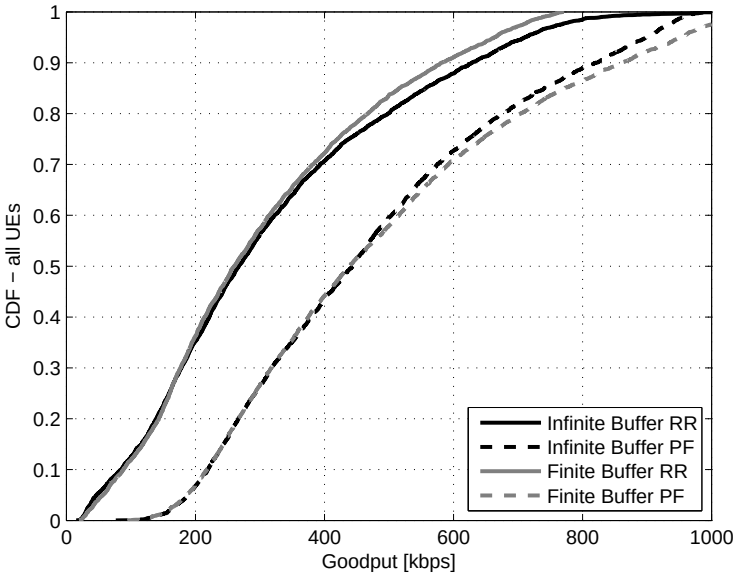

Figure 2.10: Throughput empirical Cumulative Distribution Function

- The link level AVIs have been generated by different working groups.

- As explained in appendix B, a simulations never yields a perfectly precise results. There is always a certain error margin.

2.6.5.3 Finite and infinite buffer

Figures 2.9 and 2.11 show that the system performance seems to increase with infinite buffer compared to finite buffer. In this section we clarify what are the main differences between the infinite buffer and finite buffer traffic models.

With the infinite buffer model, all UEs are staying in the network for the same amount of time. PF and RR are by nature unfair schedulers in terms of throughput. Indeed, both aim at providing the same share of resource in time and frequency domain to all UEs. As every UE has a different average channel quality, this results in unfairness in terms of throughput. This is confirmed on figure 2.11. With infinite buffer, this means that UEs with a high G-factor will download more data than UEs with low G-factor.

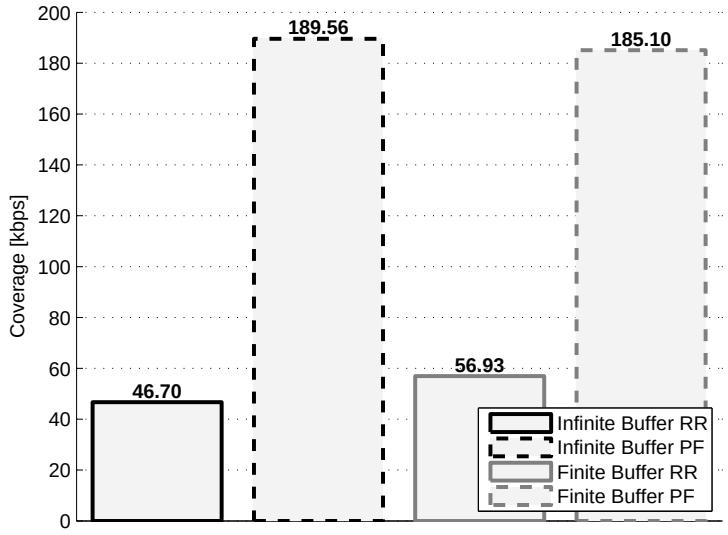

Figure 2.11: Cell Coverage: UE throughput at 5% outage

With the finite buffer model, a UE session is terminated as soon as the UE has downloaded all its buffer. As the buffer size is constant and as RR and PF are unfair, high G-factor UEs tend to stay for a shorter time in the network. This results in a general worse channel condition than with infinite buffer which explains the differences in cell throughput and coverage observed in 2.9 and 2.11.

By showing those differences, we want to highlight the importance of the traffic and call arrival modeling and the importance to take it carefully into account when interpreting results.

2.6.5.4 Round Robin and Proportional Fair

Figures 2.9 and 2.11 show as expected that PF provides a better performance in terms of coverage and cell throughput than RR. This illustrates the capacity gain brought by multi user diversity otherwise shown in various studies. However, even if PF provide a significant gain, the UE throughput depends mainly on its G-factor. In a system with QoS provision, the QoS of a UE should ideally not depend on their channel quality.

The starting point of this thesis is the performance gain brought by FDPS over non channel aware scheduling. In the following chapters, we will try to introduce various control mechanism through PS and RAC in order to give solutions for a well functioning system with QoS guarantees while translating the multi user diversity gain into QoS.

2.7 Conclusion

In this chapter, we introduced the different system assumptions and physical models used along the thesis. This framework is important to understand the future introduction of new PS and RAC algorithms. We provide also baseline results that we compared with similar studies from the literature. The comparison allows us to trust our results and validate our simulator. With the basic results, we showed the importance of the simulation methodology and assumptions. Finally, we showed that simple scheduling even including the benefits from multi user diversity gain is not sufficient for a system with QoS guarantees provisions. In the later chapters, we analyze the different PS and RAC aspect that will help providing QoS guarantees by using fully the network resources.

Chapter 3

Packet Schedulers for fairness and throughput control

3.1 Introduction

This chapter aims at introducing and analyzing new User Equipment (UE) throughput control mechanisms and principles for decoupled time / frequency domain scheduler in Orthogonal Frequency Division Multiple Access (OFDMA) systems. The proposed algorithms are evaluated in terms of coverage and cell throughput. Multi user frequency diversity gain illustrated in [23] is a key element in the algorithm design. Firstly, we briefly remind in Section 3.2 the state of the art regarding throughput control mechanisms, in Section 3.3 new controllability concepts and algorithms are introduced. Finally, the new concepts and algorithms are analyzed and discussed with simulation results in Section 3.7. Concluding remarks close the chapter in Section 3.8.

3.2 State of the Art

Many studies have been conducted in the field of Packet Scheduler (PS) for OFDMA. However, we refer here only to algorithms and methods that can be compared to the present study or being built upon by including the key characteristic of our work: low complexity. We focus therefore mainly on metric based algorithms.

PS studies for throughput control in time domain multiplexing system left many ideas that can be further applied and generalized to frequency domain multiplexing systems. For example, the time domain Proportional Fair (PF) with barrier function scheduler introduced in [33] and [37] is based on the following metric:

$$\mathbf{M}_{BF+PF}^{TD}[t][n] = \frac{\hat{\mathbf{D}}[t][n]}{\overline{\mathbf{R}}[t][n]} + \alpha \cdot e^{-\beta \cdot (\overline{\mathbf{R}} - GBR_n)} \tag{3.1}$$

where GBR_n is the Guaranteed Bit Rate (GBR) of UE n, α and β are parameters that can be set by the network operator to tune the scheduler. The scheduling metric is the sum of the PF metric and of a barrier function $\mathbf{B}_{\alpha,\beta}[t]$. The metric can therefore be rewritten in the following way:

$$\mathbf{M}_{BF+PF}^{TD}[t][n] = \mathbf{M}_{PF}^{TD}[t][n] + \mathbf{B}_{\alpha,\beta}[t][n] \tag{3.2}$$

This metric is built in a utility maximization framework thoroughly explained in [33]. In practice, PF with barrier function prioritizes UEs differently in the two different conditions:

1- if $\overline{\mathbf{R}}[t][n] << GBR_n$ then UE n does not comply with its GBR requirements. In that case, the barrier function overrides the PF metric and the lower the GBR of a UE, the higher the priority.

2- if $\overline{\mathbf{R}}[t][n] >> GBR_n$, then the PF metric prevails over the barrier function; in which case, UEs with best relative multipath constructive signal level obtain a higher priority.

In general, the priority of UEs fulfilling condition 1 is higher compared to UEs fulfilling condition 2 since the barrier function is an exponential. However, the aggressivity of the barrier function can be set thanks to the parameters α and β. The general principle of the PF with barrier function is to prioritize UEs that don't comply with their GBR requirements and at the same time to provide user diversity gain with the PF principle to the UEs that comply with their GBR. Following the same idea, the PF with Required Activity Detection (RAD) for time domain multiplexing system is introduced in [38]:

$$\mathbf{M}^{TD}_{RAD+PF}[t][n] = \frac{\hat{\mathbf{D}}[t][n]}{\overline{\mathbf{R}}[t][n]} \cdot \frac{GBR_n}{\overline{\mathbf{R_{sch}}}[t][n]} \tag{3.3}$$

Where $\overline{\mathbf{R_{sch}}}[t][n]$ describes the expected throughput when UE n is scheduled. In [38], $\overline{\mathbf{R_{sch}}}[t][n]$ is calculated by exponential filtering and corresponds to the average throughput when UE n is scheduled. The only difference being that $\overline{\mathbf{R_{sch}}}[t][n]$ is not updated when UE n is not scheduled:

$$\overline{\mathbf{R_{sch}}}[t][n] = \begin{cases} \frac{T_{sch}-1}{T_{sch}} \cdot \overline{\mathbf{R_{sch}}}[t-1][n] + \frac{1}{T_{sch}} \cdot \mathbf{R}[t-1][n] & \text{if } \mathbf{R}[t-1][n] \neq 0 \\ \overline{\mathbf{R_{sch}}}[t-1][n] & \text{if } \mathbf{R}[t-1][n] = 0 \end{cases} \tag{3.4}$$

The PF with RAD metric is the product of the PF metric and the RAD weight. It can be rewritten as follows:

$$\mathbf{M}^{TD}_{BF+PF}[t][n] = \mathbf{M}^{TD}_{PF}[t][n] \cdot \mathbf{W}[t][n] \tag{3.5}$$

The RAD weight corresponds to an estimate of the time proportion a UE should be scheduled in order to fulfill its GBR requirement. According to [38], RAD provides to each UE the required scheduling time proportion, given that the system has sufficient resource, in other words if:

$$\sum_{n=1}^{N_{UE}} \mathbf{W}[t][n] \leq (1 - BLER) \tag{3.6}$$

As it comes to OFDMA based systems, the literature has proposed different algorithms. Firstly, the Generalized Proportional Fair (GPF) scheduler has been introduced in [47]. It consists in a generalization of the PF algorithms by introducing parametric power to the nominator and denominator of the Time Domain Packet Scheduling (TDPS) and Frequency-Domain Packet Scheduling (FDPS) metrics:

$$\mathbf{M}^{TD}_{GPF}[t][n] = \frac{\hat{\mathbf{D}}[t][n]^a}{\overline{\mathbf{R}}[t][n]^b} \tag{3.7}$$

$$\mathbf{M}_{GPF}^{FD}[t][p,l] = \frac{\hat{\mathbf{d}}[t][p,l]^a}{\overline{\mathbf{R}}[t][n]^b} \tag{3.8}$$

By increasing a/b, GPF prioritizes more high G-factor UEs while by decreasing a/b, GPF tends to be more throughput fair. GPF is proved to efficiently control UE throughput fairness, however, the main drawback of this algorithms is the drop in cell capacity as a/b decreases [47]. Furthermore, GPF does not enable to control precisely the throughput in the sense that it does not specify any GBR value.

In [68] different types of TDPS and FDPS metrics are compared showing different degrees of UE throughput fairness. No control mechanism is presented in [68] but it presents new scheduling metrics and some insights on the effects of combining different types of FDPS / TDPS metrics. More precisely, the combination between the Blind Equal Throughput (BET) TDPS and Throughput To Average (TTA) FDPS provides the best throughput fairness while providing a sector throughput comparable to the PF scheduler:

$$\mathbf{M}_{BET}^{TD}[t][n] = \frac{1}{\overline{\mathbf{R}}[t][n]}. \tag{3.9}$$

$$\mathbf{M}_{TTA}^{FD}[t][p,l] = \frac{\hat{\mathbf{d}}[t][p,l]}{\hat{\mathbf{D}}[t][p]} \tag{3.10}$$

The BET TDPS aims at equalizing the past average throughputs of all UEs. Indeed, UEs with the lowest past average throughput are systematically scheduled. As long as the system is able to provide a certain share of the capacity to every UE, BET should therefore tend to equalize the past average throughput of all UEs. One of the corollary effects is that BET will tend to give a greater time fraction to low G-factor UEs than to high G-factor UEs depending on the properties of the FDPS.

TTA schedules on each Physical Resource Block (PRB) the UE with the best relative throughput. This scheduler provides multi user diversity gain as shown in [68]. However, it is characterized by its unfairness. Indeed, only a small fraction of the bandwidth is given to low G-factor UEs due to the limited range of Modulation and Coding Scheme (MCS)s as explained in [68].

In [44], [72] and [73], an entire utility based packet scheduling framework for OFDMA is introduced. The papers describe algorithms that perform optimum scheduling with the gradient algorithm. the optimization is based on concave utility functions depending on user data rate or packet delivery delay. Those studies propose therefore an alternative to the simple metric based algorithm presented in Chapter 2 but with increased complexity. In the PhD study, only the metric based algorithms with a decoupled time and frequency packet scheduler are studied as it provides a guarantee for low complexity.

3.3 Throughput Controllability: Principle and Algorithms

In this section, we introduce a general method for throughput fairness controllability as illustrated in Figure 3.1. Firstly, in our PS method, we aim at providing always an optimal multi user diversity gain. The multi user diversity gain increases with the number of UEs scheduled at the

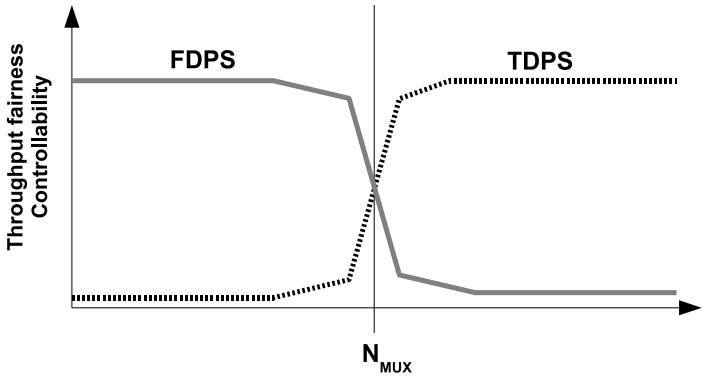

Figure 3.1: Throughput fairness controllability principle

same time. Therefore, the TDPS is designed to always pass the maximum number of UEs to the FDPS: $min\,(N_{MUX}, N_{UEs})$. From a throughput fairness controllability perspective, the PS can meet two situations:

$N_{UE} \ll N_{MUX}$*:* We define the time fraction TF as the proportion of time where a UE is scheduled. The total available time fraction in the system is exactly N_{MUX}. Indeed, as a maximum of N_{MUX} UEs can be scheduled at the same time, up to N_{MUX} UEs can be scheduled every Transmission Time Interval (TTI). The maximum time fraction of a UE is of course 1 and the minimum is 0. The UE average time fraction $\overline{TF}_{user}$ is defined as the time fraction given to each UE by a time fair TDPS:

$$\overline{TF}_{user} = \frac{N_{MUX}}{N_{UE}} \tag{3.11}$$

Of course, as the number of UEs increases, the UE average time fraction decreases. At the same time, the relative time fraction allocated to each UE can be modulated with more flexibility when the number of UEs increases. Another aspect of the TDPS allocation flexibility is the number of allocation combination $C_{N_{UE}}^{N_{MUX}}$ that increases with N_{UE}:

$$C_{N_{UE}}^{N_{MUX}} = \frac{N_{UE}!}{N_{MUX}!\,(N_{UE} - N_{MUX})!} \tag{3.12}$$

We want to stress here the increased impact of the TDPS with a high N_{UE}. As such a situation allows a good control of the UE relative time fraction, we propose a strategy where the TDPS is the sole responsible for the throughput control through control of the time fraction allocated to each UE while the FDPS is purely a multi user diversity gain mechanism.

$N_{UE} \lesssim N_{MUX}$*:* When the number of UEs becomes smaller, the influence of the TDPS decreases until it is completely deactivated when $N_{UE} < N_{MUX}$. In that case, the time fraction given to each UE is simply equal to 1. As the TDPS has low impact or is deactivated, the FDPS must be used for both purpose: throughput fairness control and multi user diversity gain.

3.4 Throughput Control with TDPS: the Priority Set Scheduler

We introduce here a new TDPS algorithm inspired by the time domain PF with barrier function. It consists in giving a high priority to the UEs that do not comply with their GBR and a lower priority

to those that comply with their GBR. Formally, Priority Set Scheduler (PSS) consists in separating UEs in two distinct sets with strict prioritizing of the UEs in set 1 over the UEs in set 2. The prioritizing within each set is performed with a metric specific to each set. The set differentiation criterions of PSS are described in Table 3.1.

Table 3.1: Priority Set Scheduler

Set	criterion UE n	Internal set metric
1	$\overline{\mathbf{R}}[t][n] \leq GBR_n$	normalized BET (3.13)
2	$\overline{\mathbf{R}}[t][n] > GBR_n$	normalized PF (3.14)

UEs of set 1 are given absolute priority over UEs of set 2. Therefore, .The metrics used for the PSS algorithms are GBR normalized versions of BET and PF. Normalized BET prioritizes UEs within set one. It prioritizes UEs with the highest relative distance to their GBR:

$$\mathbf{M}_{NBET}^{TD}[t][n] = \frac{GBR_n}{\overline{\mathbf{R}}[t][n]}. \tag{3.13}$$

while normalized PF prioritizes UEs within set two:

$$\mathbf{M}_{NPF}^{TD}[t][n] = GBR_n \cdot \frac{\hat{\mathbf{D}}[t][n]}{\overline{\mathbf{R}}[t][n]} \tag{3.14}$$

The normalizations in the BET and PF metrics (equations 3.13 and 3.13) aims at treating fairly UEs with different GBR values. Generally, PSS is a simplification of the PF with barrier function algorithm expressed in (3.1). PF with barrier function requires the settings of parameters α and β, supposed to help managing different levels of priority between different UEs. PSS is a simplification in the sense that it does not include any differentiation mechanism. However, no setting is required at all in PSS which makes it directly functional.

3.5 Decoupling between TDPS and FDPS

If the TDPS takes care of modulating the time fraction given to each UE in order to regulate the UE throughput, the FDPS should not include any throughput control mechanism that is in contradiction with the TDPS. For example, FDPS-PF described by the metric in (2.16) includes a control mechanism that can be in contradiction with PSS. In order to illustrate the contradiction, let's consider the following simple situation where all UEs have the same GBR:

$$\forall \left(n, n'\right) \in [1, N_{UE}]^2 : \quad GBR_n = GBR_{n'} = GBR \tag{3.15}$$

PSS can be seen as a modification of TDPS-PF that aims at increasing the time fraction of lower G-factor UEs so that they can be provided their GBR. In other words, a set of UEs $S_{PF}^{<GBR}$ that could not be provided their GBR with a simple TDPS-PF scheduler will regularly fall into set 1 in PSS so that on the long term:

$$\text{if } n \in S_{PF}^{<GBR} : \quad \overline{\mathbf{R}}[t][n] \approx GBR_n \tag{3.16}$$

Note that this is the case only if the GBR requirements of the system are not too high. With condition specified (3.15), the consequence is that UEs of set $S_{PF}^{<GBR}$ will have a FDPS-PF metric with an approximately constant denominator:

$$\text{if } n \in S_{PF}^{<GBR}: \quad \mathbf{M}_{PF}^{FD}[t][n,l] = \frac{\hat{\mathbf{d}}[t][n,l]}{\overline{\mathbf{R}}[t][n]} \approx \frac{\hat{\mathbf{d}}[t][n,l]}{GBR} \tag{3.17}$$

This means that within $S_{PF}^{<GBR}$, UEs with higher G-factor will be prioritized over UEs with a lower G-factor. This behavior is not consistent with the PSS principle as PSS uses the BET metric for UEs that fall into set 1 and as explained in section 3.2, BET prioritizes UEs with low G-factor.

This example illustrates the need for FDPS algorithms that do not include any independent throughput control mechanism. We introduce therefore to new FDPS algorithms, which sole purpose is to provide multi user diversity gain.

3.5.1 Carrier over Interference to Average

The Carrier over Interference to Average (CoItA) metric is defined as as the ratio between an estimate of the Signal-to-Interference-plus-Noise Ratio (SINR) on PRB k and an estimate of the G-factor:

$$\mathbf{M}_{CoItA}^{FD}[t][p,l] = \frac{\hat{\mathbf{SINR}}[t][p,l]}{\hat{\mathbf{G}}[p]} \tag{3.18}$$

Under the following conditions:

- *Condition 1* All E-UTRAN Node B (eNode-B)s transmit with full power
- *Condition 2* The different interferers are uncorrelated and are in sufficient number so that the departure from average of the denominator of the SINR in (2.2) is not significant compared to the departure from average of the numerator.

it is possible to approximate the SINR as:

$$\mathbf{SINR}[t][p,k] \approx \mathbf{G}[p] \cdot \sum_{a=1}^{N_{Rx}} \mathbf{HH}^{i}[a][p,k] \tag{3.19}$$

Note that the accuracy of *Condition 2* depends on the physical conditions of the considered UE. For example, if the UE is close to the eNode-B, the departure from average of the SINR denominator will be naturally small compared to that of the numerator. Indeed, the various main interferers will be placed at equivalent distance of the UE thus having a similar path gain. They will therefore create a cancellation effect as the different interferers fade independently. However, if a UE is far from the eNode-B, at cell edge, their may be an interferer providing an interference level significantly higher than all the others. In that case, the departure from average of the SINR denominator may be high and the approximation of (3.19) may be less accurate.

By combining (3.19) and (3.18), we can write:

$$\mathbf{M}_{CoItA}^{FD}[t][p,l] \approx \sum_{a=1}^{N_{Rx}} \mathbf{HH}^{S}[a][p,l] \tag{3.20}$$

This means that CoItA schedules on a given PRB the UE which is in the highest fade. Note that $\sum_{a=1}^{N_{Rx}} \mathbf{HH}^i[a][p,l]$ is central chi-square distributed with $2 \cdot N_{Rx}$ degrees of freedom and as each link fades independently, with Rayleigh fading, $\sum_{a=1}^{N_{Rx}} \mathbf{HH}^i[a][p,l]$ are Independent and Identically Distributed (i.i.d) among all UEs which means that all UEs have equal probability of being scheduled. The consequence is that the average number of PRBs allocated to each UE when selected by the TDPS is approximately equal. We can summary the properties of CoItA:

- on a given PRB, CoItA schedules the UE in the highest fade;
- CoItA provides equal quantities of PRBs among UEs.

As the definition of CoItA in (3.18) is only a theoretical form, we need to provide a definition that uses only elements that the eNode-B has access to in agreement with the system description in Chapter 2:

$$\mathbf{M}_{CoItA}^{FD}[t][p,l] = \frac{\mathbf{CQI}^{trans}[t][p,l]}{\sum_{l'=1}^{N_{PRB}} \mathbf{CQI}^{trans}[t]\left[p,l'\right]} \tag{3.21}$$

(3.21) uses mainly the Channel Quality Information (CQI) defined in (2.8) as an SINR estimate and the sum of the CQI as a G-factor estimation.

3.5.2 Proportional Fair scheduled

We introduce here a modified version of the FDPS-PF algorithm. The goal of this modified version called Proportional Fair scheduled (PFsch) is to remove the contradictory behavior of FDPS-PF described in section 3.4. The PFsch scheduler is defined by the following metric:

$$\mathbf{M}_{PF_{sch}}^{FD}[t][p,l] = \frac{\hat{\mathbf{d}}[t][p,l]}{\overline{\mathbf{R_{sch}}}[t][n]} \tag{3.22}$$

where

$$\overline{\mathbf{R_{sch}}}[t][n] = \begin{cases} \frac{T_{sch}-1}{T_{sch}} \cdot \overline{\mathbf{R_{sch}}}[t-1][n] + \frac{1}{T_{sch}} \cdot \mathbf{R}[t-1][n] & \text{if } \mathbf{R}[t-1][n] \neq 0 \\ \overline{\mathbf{R_{sch}}}[t-1][n] & \begin{array}{l}\text{if UE } n \text{ is not}\\ \text{scheduled by}\\ \text{TDPS at}\\ \text{TTI } t-1\end{array} \end{cases} \tag{3.23}$$

Note that the definition of $\overline{\mathbf{R_{sch}}}$ is slightly different in (3.23) and (3.4). in (3.23), $\overline{\mathbf{R_{sch}}}$ is not updated when the UE is not scheduled by the TDPS while, in (3.4) $\overline{\mathbf{R_{sch}}}$ is not updated when the throughput of the UE is equal to 0 in the previous TTI. The difference lies in the fact that there can be a situation were a UE is scheduled by the TDPS but not allocated to any PRB by the FDPS. By introducing the modification in the definition, $\overline{\mathbf{R_{sch}}}[t][n]$ can be interpreted as the average throughput when UE n is scheduled by the TDPS.

Contrarily to FDPS-PF, FDPS-PFsch does not involve a control mechanism that controls the UE throughput. FDPS-PFsch involves the use of $\overline{\mathbf{R_{sch}}}$ which is updated only when the concerned UE is scheduled by the TDPS thus decoupling completely the TDPS and the FDPS: scheduling strategy of the TDPS will have a limited influence on the FDPS.

3.6 Throughput Control with Frequency-Domain Packet Scheduling

When the number of UEs in the system is lower than N_{UE}, no TDPS is applied as explained in section 3.3 and therefore, the throughput control through TDPS is deactivated. We therefore introduce a new throughput control mechanism through FDPS. Our control mechanism is inspired of the RAD principle and consists in applying a weight to the FDPS metric depending on the throughput requirements of the UE. The weighted FDPS metric can be described as follows:

$$\mathbf{M}^{FD}_{X+W}[t][p,l] = \mathbf{W}^{FD}[t][p] \cdot \mathbf{M}^{FD}_{X}[t][p,l] \tag{3.24}$$

where X is one of the FDPS metrics like PF, TTA described in section 3.2 or PFsch, CoItA described in section 3.7.2. $\mathbf{W}^{FD}$ is the FDPS weight defined as:

$$\mathbf{W}^{FD}[t][p] = \max\left(1, \frac{GBR_p}{\overline{\mathbf{R_{sch}}}[t][p]}\right) \tag{3.25}$$

The FDPS weight aims at increasing the number of PRBs given to a UE that does not comply with its GBR by increasing the priority given to this UE. At the same time, UEs that comply with their GBR requirement have a weight equal to 1.

3.7 Performance Evaluation

In this section, we analyze the different algorithms and concepts proposed in this chapter. Firstly, we evaluate the performance of the TDPS-PSS algorithm coupled with different FDPS algorithms. Then, in a second part, we show the performance of throughput control with FDPS weighting.

3.7.1 Throughput Control with TDPS

In order to assess the performance of TDPS-PSS and illustrate the influence between throughput control with TDPS and different FDPS algorithms, simulations have been run varying the following two factors:

- *the FDPS algorithms*: PSS is evaluated with four different FDPS algorithms. Two algorithms from the literature: PF and TTA. Two algorithms proposed in this chapter: PFsch and CoItA.
- *the GBR*: In the simulations, every UE has the same GBR requirement, following the simplified assumption in (3.15). The algorithms are evaluated with GBR values varying between 0kbps and 400kbps. The chosen values span over a range that explores all possible states of the system in terms of coverage and average cell throughput.

The detailed simulations assumptions are summarized in table 3.2.

Figure 3.2 shows the GBR values versus coverage and average cell throughput for the different FDPS algorithms. The general tendency is that the coverage increases with the GBR while the average cell throughput decreases. Generally, by modulating the GBR, the TDPS-PSS algorithm

Table 3.2: Throughput control with TDPS: simulation parameters

Parameter	**Setting**
Envirronment	MACRO #1
Layout Configuration	1 simulated site
Traffic models / call arrival	Infinite Buffer N_{run}=50 T_{IB}=10s N_{UE}=30
TDPS	PSS GBR=0kbps =200bkps =250bkps =300bkps =400bkps
FDPS	TTA PF PFsch CoItA

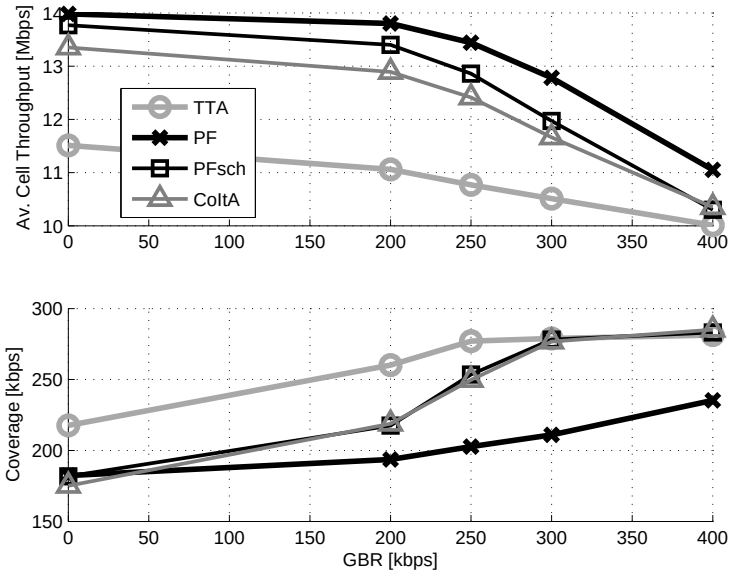

Figure 3.2: Performance in terms of average cell throughput and coverage of TDPS-PSS with different FDPS evaluated for different values of GBR.

enables to trade cell throughput for coverage. We therefore show here that TDPS-PSS is a mean to control throughput fairness in a situation where $N_{UE} \ll N_{MUX}$.

However, we can observe that the different FDPS algorithms have a different influence on the flexibility of the cell throughput / coverage trade. We observe three different behaviors that we can characterize as follows:

- *TTA* generally shows, compared to other FDPS algorithms, a higher coverage and a lower average cell throughput. The coverage varies between 217kbps and 281kbps, thus spans over a small range of 64kbps.

- *PF* show, compared to other FDPS algorithms, a lower coverage and a higher average cell throughput. The coverage also varies over a small range of 53kbps between 182kbps and 235kbps.

- *PFsch and CoItA* show a wide span of coverage between 175kbps and 285kbps. Both coverage and average cell throughput cover a wide range of values.

Finally, the best throughput fairness controllability is obtained with TDPS-PSS combined with CoItA or PFsch as those combinations offer the greatest flexibility in terms of cell throughput / coverage trade-off. We can distinguish three different zones. Between, $GBR = 0$ and $GBR = 200$kbps, the coverage increases little and the average cell throughput is nearly constant. For those GBR values, only few UEs fall into set 1 of the PSS algorithms and the PF part of PSS guarantees by itself a throughput higher than the preset GBR. Between, $GBR = 200$kbps and $GBR = 300$kbps, the scheduler is in a zone that we will call the *working zone*. Between those values, the GBR is very close to the coverage value, therefore in that zone, the algorithm is reaching its target. Finally, $GBR > 300$kbps, the coverage is constant and the average cell throughput decreases with the GBR. We call that zone, the *saturation zone*. In that zone, UEs with a throughput below the 5^{th} percentile are given the maximum throughput that can be provided. By increasing the GBR, PSS tries to equalize the throughput of all UEs and tends toward the BET scheduler as it can provide to only a very small proportion of the UE the required GBR. Figure 3.3 confirms it and shows that for $GBR = 300$kbps, around 20% of the UEs are bellow 300kbps while for $GBR = 400$kbps, 95% of the UEs are bellow 400kbps. the saturation zone is not desirable as it does not provide Quality of Service (QoS) to any UE and decreases the average cell throughput. The regular mechanism to avoid falling into saturation zone is the Radio Admission Control (RAC) studied in a later stage in this thesis.

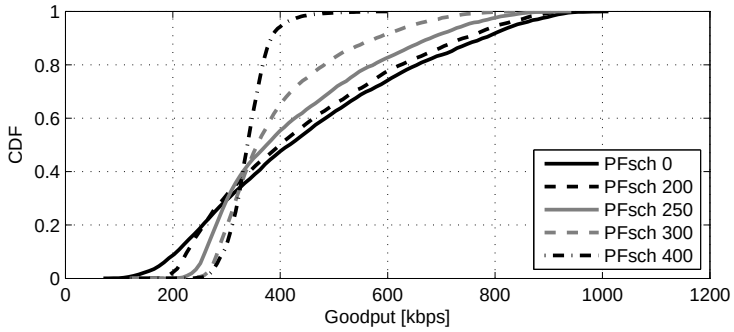

Figure 3.3: UE throughput CDF for TDPS-PSS FDPS-PFsch with different GBR values.

3.7.2 Decoupling between TDPS and FDPS

In order to explain the different tendencies of the different combinations between PSS and FDPS algorithms, we need to understand the characteristic behaviors of the different FDPS algorithms. To this end, figure 3.4 shows the average number of PRBs scheduled per UE when scheduled by the TDPS versus different G-factor values. From now on, we call this graph the *PRB profile* of a scheduler.

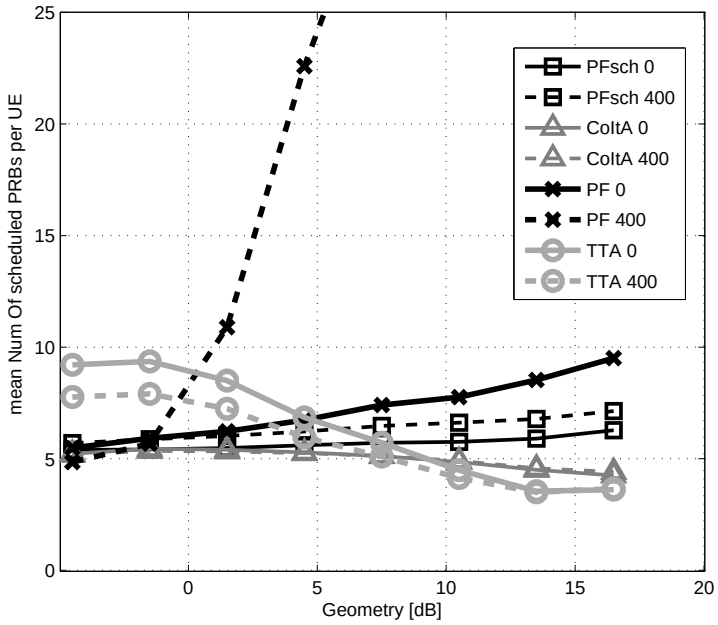

Figure 3.4: G-factor versus average number of scheduled PRBs per UE. Values are given for different FDPS algorithms with different GBR values.

We firstly observe that the PRB profile of PF changes drastically depending on the GBR value used by PSS. for $GBR = 0$kbps, the average number of PRB scheduled increases with the G-factor from 5 to 10, while for $GBR = 400$kbps the average number of PRB increases from 5 to more than 25. This is the direct consequence of the principle explained in section 3.7.2. By increasing the GBR, the number of UEs that fall into set $S_{PF}^{<GBR}$ increases. Therefore, the priority given to high G-factor UEs is more visible on the PRB profile when GBR increases. Finally, while PSS tries to give a more important time fraction to low G-factor UEs, FDPS-PF does the total opposite. This contradictory behavior has two consequences. Firstly, it limits the controllability of the throughput as seen on figure 3.2. Secondly, it limits the benefit of multi user diversity as to large parts of the spectrum are allocated to a unique UE. This effect is shown later in this section.

The PRB profile of TTA show an average number of scheduled PRBs that increases from -8dB to -3dB and then decreases. The very small difference between the two PRB profiles depending on the GBR indicates that the TDPS-PSS and FDPS-TTA are decoupled and do not interact with each other. The decreasing slope of the PRB profile of TTA can be partly explained by a limitation of the Link Adaptation (LA) functionality. Indeed, the different throughput estimations made by the LA functionality are done considering only the range of available MCS. Therefore, the number of possible throughput estimates on a PRB is equal to the number of MCS available in the system. The MCS are available on a certain range that does not exactly overlap the SINR range of the cell.

Therefore, a UE which is close to the eNode-B might be in a situation where its SINR is often so high that there is no MCS that can match this SINR on a given PRB. Therefore, in that case, the estimated throughput may be underestimated and therefore the TTA metric is also lower than it would be if the system could support higher SINRs. The exact opposite effect also occurs for very low SINRs where in that case, the throughput is overestimated. These effects help explain why more PRBs are allocated to low Geometry UEs.

The PRB profile of PFsch is similar to that of PF with PSS and $GBR = 0$kbps. However, the PRB profile of PFsch changes only very little with the GBR settings. PFsch reaches therefore clearly its goal of providing the advantages of PF in terms of multi user diversity gain and at the same time being independent of the TDPS.

Finally, the PRB profile of CoItA is nearly flat as expected from the explanations in section 3.5.1

Figure 3.5 is a representation of the PS performance. We want to compare the different schedulers in terms of combined performance cell throughput / coverage. Note that Figure 3.5 uses the same data than figure 3.2. With infinite buffer simulations, it is capital to show both cell throughput and coverage performance on the same graph in order to compare the overall performance.

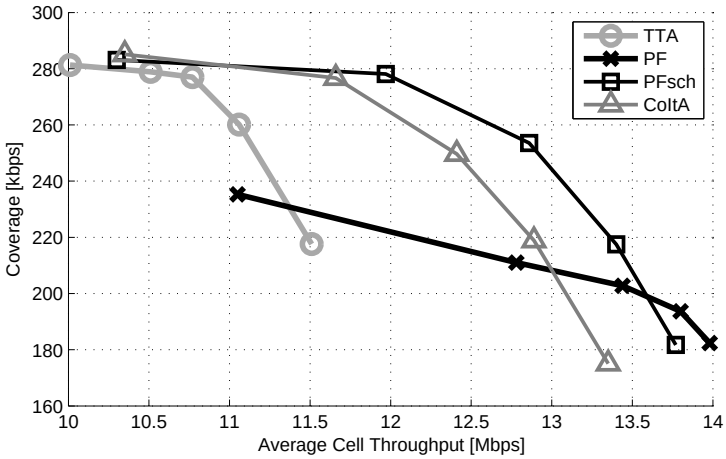

Figure 3.5: Average cell throughput versus coverage for TDPS-PSS with different FDPS algorithms.

We can observe that for the same coverage values, TTA provides a cell throughput very inferior to CoItA and PFsch. The situation is similar for PF. We show here that PFsch and CoItA provide a better resource utilization than TTA and PF due to a better exploitation of the multi user diversity.

3.7.3 Throughput Control with FDPS

In order to evaluate FDPS weighting, we run simulations with 10 UEs per cell and therefore the TDPS is deactivated. This type of setting allows us to analyze in details the control capabilities of FDPS weighting. The detailed assumptions are summarized in table 3.3. Similarly to the TDPS control simulations, we vary the following parameters:

- *the FDPS algorithms*: we evaluate PSS with three different FDPS algorithms: TTA, PF and CoItA. As we run simulations with the TDPS deactivated, PF and PFsch are equivalent.

Table 3.3: Throughput control with TDPS: simulation parameters

Parameter	**Setting**
Envirronment	MACRO #1
Layout Configuration	1 simulated site
Traffic models / call arrival	Infinite Buffer N_{run}=150 T_{IB}=10s N_{UE}=10
FDPS weight	PSS GBR=400kbps =500bkps =700bkps =900bkps =1100bkps
FDPS	PFsch CoItA

- *the GBR*: In our simulations, every UE has the same GBR requirement, following the simplified assumption in (3.15). The algorithms are evaluated with GBR values varying between 400kbps and 1500kbps. The chosen values span over a range that explores all possible states of the system in terms of coverage and average cell throughput. Note that the range spans over higher values than for TDPS control as with less UEs, higher coverage is expected.

Figure 3.6 shows that FDPS weighting generally allows to trade cell throughput for coverage. As for TDPS-PSS we can observe different types of coverage and average cell throughput spans and controllability depending on the FDPS algorithms used:

- TTA has a nearly constant cell throughput of around 10.5Mbps while the coverage varies between 600bkps and 783kbps. TTA helps reaching the highest coverage.
- PF has the highest cell throughput that decreases from 13.7Mbps down to 12.4Mbps with a coverage that increases up to 674kbps.
- CoItA has the largest span in terms of coverage: from 536kbps to 780kbps. Together with TTA, CoItA helps reaching the highest coverage. However, the average cell throughput is generally lower and goes from 13.06Mbps down to 11.25Mbps.

Firstly, FDPS weighting contrasts with TDPS-PSS in the sense that it is not possible to identify any *working zone*. Indeed, generally, the coverage value increases slower than the GBR. Figure 3.7 shows the different UE throughput CDFs for CoItA. We can observe that FDPS weighting is not aggressive enough to provide their GBR to all UEs.

The differences of control flexibility with the different FDPS algorithms can be explained by observing the PRB profiles of the different FDPS algorithms with different GBR settings on figure 3.8. When the GBR is set to a high value, FDPS weighting changes the PRB profile. Generally,

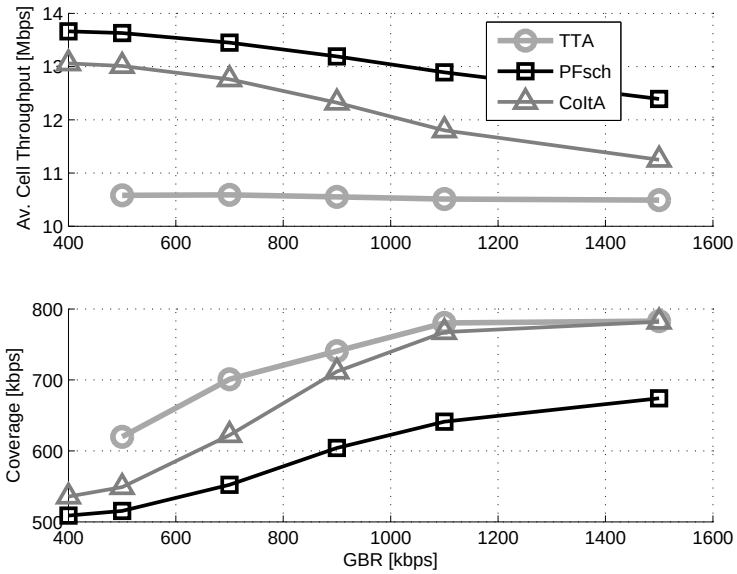

Figure 3.6: Performance in terms of average cell throughput and coverage of FDPS-weighting with different FDPS evaluated for different values of GBR.

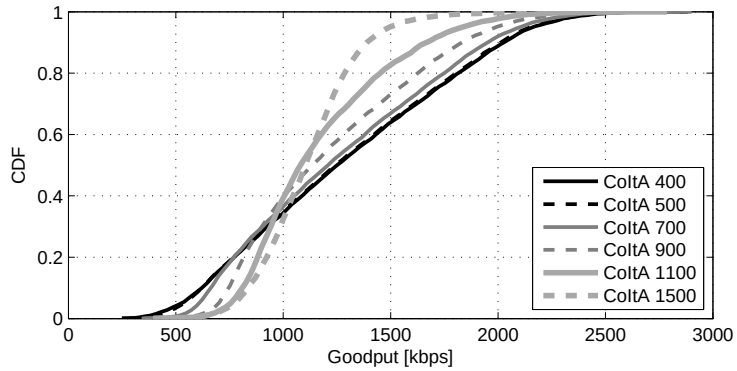

Figure 3.7: UE throughput CDF for FDPS weighting with FDPS-CoItA with different GBR values.

more PRBs are allocated to lower G-factor UEs when the GBR increases, which is the actual goal of FDPS weighting. However, as explained in Section 3.7.2 and as shown in Figure 3.8, FDPS-TTA tends naturally to give more PRBs to low G-factor UEs. Therefore, the effect of weighting is very limited and the controllable span of coverage and cell throughput is therefore lower. Note on figure 3.8 the small difference between the PRB profiles of FDPS-TTA with GBRs of 500kbps and 1500kbps. FDPS-CoItA and FDPS-PF have however naturally a flatter PRB profile. This allows those two algorithms to propose wider possibilities for throughput-coverage trade-off.

Finally, Figure 3.9 show coverage versus average cell throughput for the different FDPS algorithms. We see clearly that FDPS-PF and FDPS-CoItA outperform FDPS-TTA in terms of coverage and cell throughput.

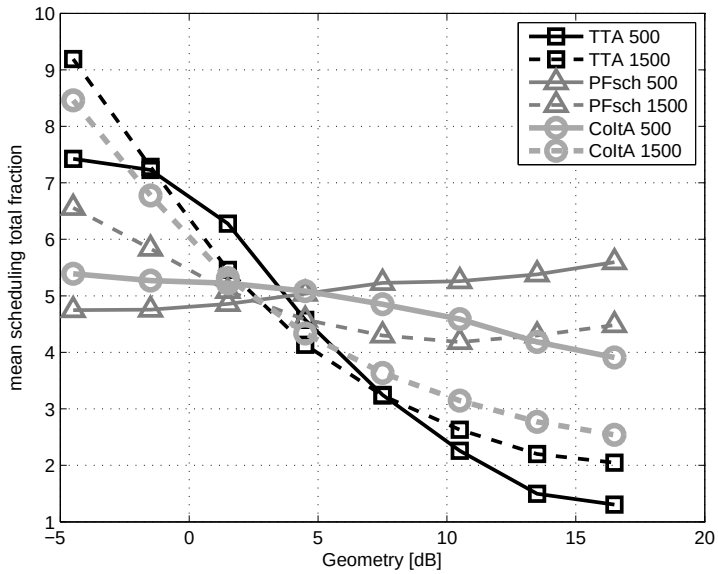

Figure 3.8: G-factor versus average number of scheduled PRBs per UE. Values are given for different FDPS algorithms with different GBR values.

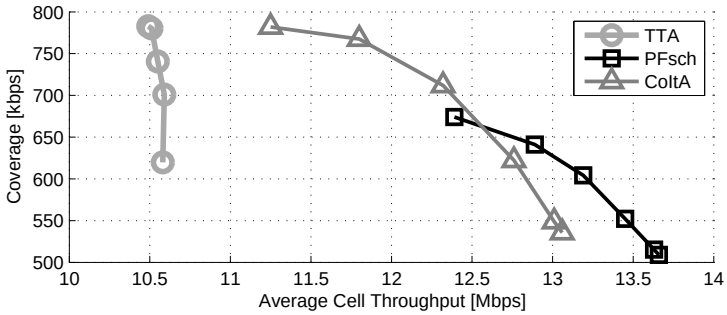

Figure 3.9: Average cell throughput versus coverage for FDPS-weighting with different FDPS algorithms.

3.8 Conclusion

Throughput control with the decoupled time and frequency domain packet scheduler had not been studied in any publication (to the knowledge of the author). In this Chapter two general design principles for performing throughput control with a decoupled packet scheduler are highlighted.

Firstly, it is possible to control the UE throughput with different packet scheduler entities depending on the number of UEs present in the system. When the number of UEs is large compared to N_{MUX} , the throughput can be controlled with the TDPS. For example PSS is a very simple algorithm that can perform that task. In the opposite case, the throughput can be controlled by the FDPS through FDPS-weighting. Locating the throughput control functionality only in one component of the packet scheduler allows to benefit fully from multi user diversity gain.

Secondly, the TDPS and the FDPS should be independent. More specifically, it means that the FDPS should not include any control mechanism that is contradictory with the TDPS. The two FDPS algorithms introduced CoItA and PFsch do not include any throughput control mechanism and are therefore independent from the TDPS. Furthermore, they are the best solutions as they

provide the best multi user diversity gain and the best flexibility for the throughput control methods. Indeed, at equal coverage, CoItA and PFsch provide cell throughput gains of the order of 10% over the reference algorithms TTA.

Chapter 4

Packet Scheduling Under Fractional Load Conditions

4.1 Introduction

We have assumed so far in the thesis that the E-UTRAN Node B (eNode-B) is transmitting with a constant power spectral density over the whole transmission bandwidth. In this chapter, we introduce fractional load scenarios. We define fractional load by a situation where:

- the eNode-B transmits only over a fraction of the bandwidth,
- the used bandwidth fraction can change from Transmission Time Interval (TTI) to TTI.

Following the thesis framework and the assumptions presented in Section 2.2. A Physical Resource Block (PRB) can be either in 'off' mode where it is not transmitting any data or in 'transmission' mode. When a PRB is in 'transmission' mode, it transmits always with the same power. A fractional load scenario is therefore a two-state transmit power variations in frequency and time domain thus creating interference variations in neighboring cells. Power variations can have two effects on a downlink Orthogonal Frequency Division Multiple Access (OFDMA) system.

- *Negative effect:* Too fast interference variation can compromise the relevance of the Channel Quality Information (CQI) reports as the CQI report is delivered with a delay.
- *Positive effect:* A partial use of the bandwidth in a cell means better Signal-to-Interference-plus-Noise Ratio (SINR) conditions in the neighboring cells.

Those effects are handled by techniques usually called: Inter-Cell Interference Coordination (ICIC). In section 4.2, we propose ICIC techniques, which we call PRB pattern selection as they consists in selecting a certain transmitting PRB pattern every TTI with the goal of minimizing the reporting delay effects and maximizing the global SINR conditions. Furthermore, the proposed algorithms are autonomous as they are based only on the information available in the own cell and therefore do not require any inter-eNode-B signaling over the X2 interface.

The performance of the proposed PRB pattern selection algorithms is first studied under different loadings in terms of coverage and average cell throughput. In the rest of the chapter, we try to put fractional load scenarios into a realistic context.

Firstly, fractional load can be seen as a situation voluntarily created in order to decrease interference and increase the system coverage. In section 4.3, we study fractional load as a coverage enhancement technique. The choice of the Packet Scheduler (PS) algorithms is seen here as a critical element and a strong emphasis is put on that aspect.

Then, fractional load can simply be a situation that occurs by itself in case of low traffic load. Indeed, the Frequency-Domain Packet Scheduling (FDPS) metric based algorithm described in section 2.6.3 does not guarantee a full usage of the bandwidth. In that case, the prime goal of PRB pattern selection is to minimize the bandwidth used in order to improve the SINR conditions in neighboring cells. In section 4.4, we study fractional load in realistic low traffic condition and soften the PRB pattern selection concept into PRB pattern prioritizing.

Finally, we close the chapter with concluding remarks in Section 4.5

4.2 PRB pattern selection

The integration of PRB pattern selection to the PS framework is illustrated in figure 4.1. The elements represented with dashed lines are new compared to the basic PS framework in figure 2.6. The PRB pattern can be chosen thanks to several input parameters:

- *The Bandwidth Fraction Factor (BFF)* is the fraction of the bandwidth to be used by the PRB pattern. The BFF determines the number of transmitting PRBs N_{PRB}^{Tx} as:

$$N_{PRB}^{Tx} = N_{PRB} \cdot BFF \tag{4.1}$$

- *The CQI* (as SINR estimation or PRB throughput estimation from Link Adaptation (LA)) can be used by the PRB pattern selection in order to choose the PRBs with the least interference.
- *The previous PRB patterns* can be used to monitor and control the changes in time of the PRB pattern.

The PRB pattern selection is integrated in the matrix based FDPS algorithm described in Section 2.6.3 with two modifications. The modifications are located in the second step of the algorithms entitled *Allocate UEs without retransmission*. In the version described in Section 2.6.3:

- The initial set of schedulable PRBs S_{sch}^{PRB} consists of all the PRBs.
- The conditional argument of the while loop is $N_{schPRB} < N_{PRB} - x$

In, the modified version, for integration of PRB pattern selection, the following modifications are applied:

- The initial set of schedulable PRBs S_{sch}^{PRB} consists of the PRB pattern passed by the PRB pattern selection module.

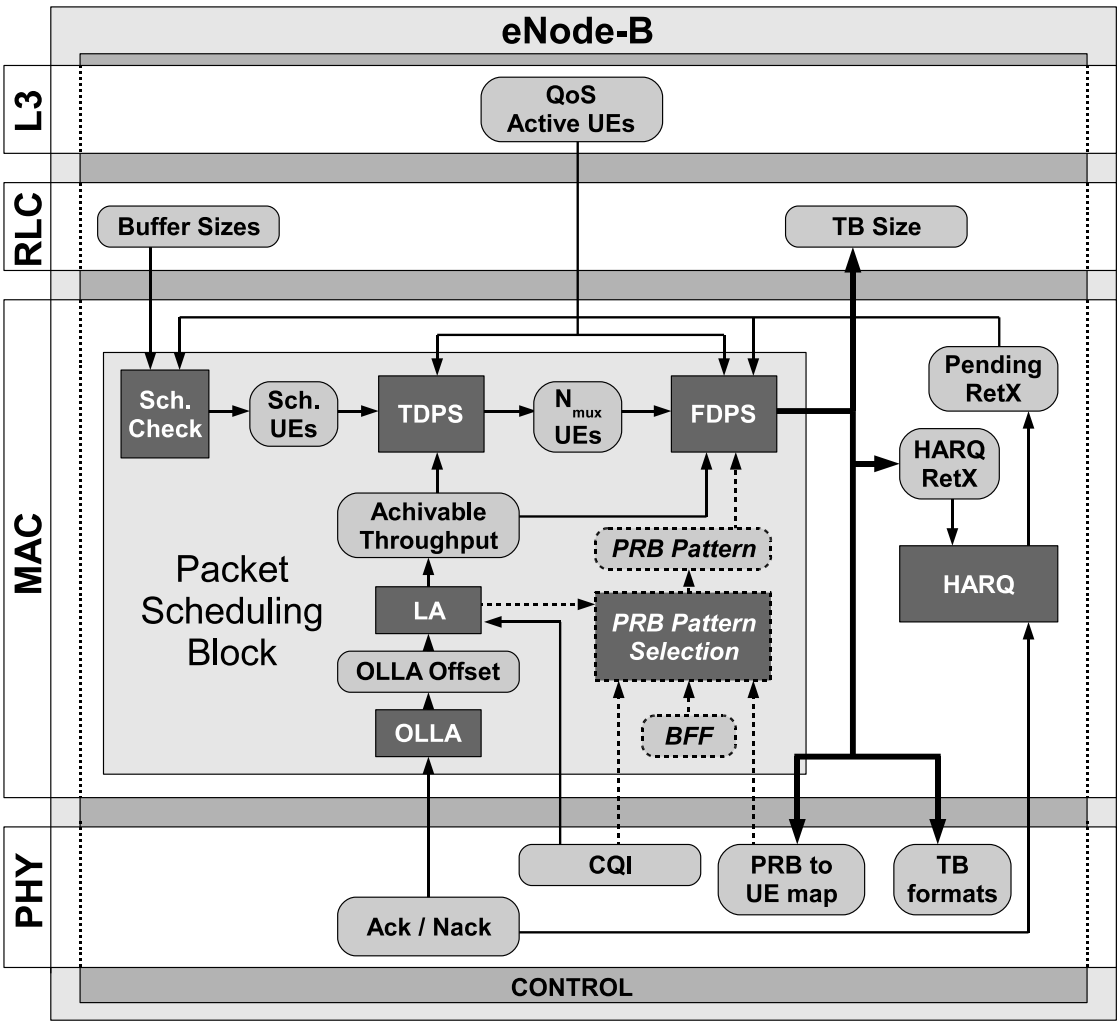

Figure 4.1: Conceptual integration of PRB pattern selection to the PS Model. New elements are represented with dashed lines.

- The conditional argument of the while loop is $N_{schPRB} < N_{PRB}^{Tx} - x$

In the rest of the section, we introduce different PRB pattern selection methods.

4.2.1 Best Metric

The Best Metric (BM) PRB pattern selection algorithm follows the same method than in [74]. It is the most simple algorithms presented in this study. It inherits directly from the FDPS. At TTI t It selects the N_{PRB}^{Tx} PRBs with the indices l that maximize the following value:

$$\mathbf{Q}_{BM}\left[t\right]\left[l\right] = \max_{p} \mathbf{M}^{FD}\left[t\right]\left[p,l\right] \tag{4.2}$$

Where p is the index for the User Equipment (UE)s and $\mathbf{M}^{FD}$ is the FDPS metric. In a situation with non buffer limited UEs, $\mathbf{Q}_{BM}[k,t]$ corresponds to the metric value of the UE scheduled on PRB l as described in section 2.6.3. BP is FDPS implicit in the sense that it relies on the properties of the FDPS algorithms to create ICI mitigation.

4.2.2 Random Correlated PRB Pattern

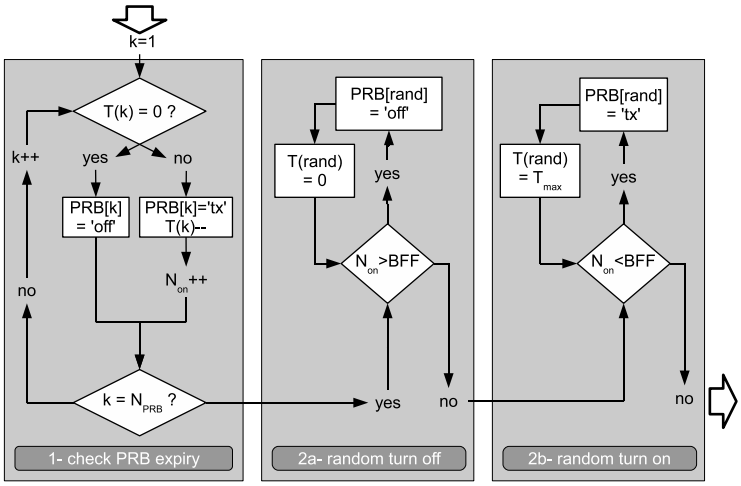

Figure 4.2: Description of the Random Correlated PRB Pattern PRB pattern selection algorithm.

Random Correlated PRB Pattern (RCPP) is described fully on Figure 4.2. It aims at forcing time correlation in the used PRB pattern. Time correlation is achieved by setting a counter on a PRB when it is turned into the 'transmission' (abbreviated 'tx' in the drawing) state. The counter is initialized at the value T_{Max} which corresponds to the maximum number of consecutive sub-frames where a PRB can be in the 'transmission' state. The PRB is automatically turned in the 'off' state when the counter reaches 0. When the PRB pattern includes less than N_{PRB}^{Tx} PRBs, PRBs are turned in the 'transmission' state randomly among the PRB in the 'off' state. Inversely, when the PRB pattern includes less than N_{PRB}^{Tx}, PRBs are turned 'off' randomly among the PRBs in 'transmission' mode. Note that RCPP is not channel aware; The only purpose of RCPP compared to BM is to add time correlation in the used PRB pattern in order to avoid to fast variations of the channel and to increase the relevance of the CQI.

4.2.3 Best Quality PRB Pattern

Best Quality PRB Pattern (BQPP) aims at selecting the PRBs with the least average interference while forcing the time steadiness of the PRB pattern. This strategy follows a mutual benefit principle. Choosing PRBs with the least average interference in the own cell means also choosing the PRBs that are the least in 'transmission' state in neighboring cells which implicitly means that the own cells tries also to minimize the neighboring cells average interference. Practically, BQPP selects the N_{PRB}^{Tx} PRBs that maximize the quality metric. The quality metric of PRB l at sub-frame t: $\mathbf{Q}_{BQPP}[t][l]$ is defined by an exponential time averaging of the instant quality metric $\mathbf{IQ}[t][l]$:

$$\mathbf{Q}_{BQPP}[t][l] = \frac{T-1}{T}\mathbf{Q}[t-1][l] + \frac{1}{T}\mathbf{IQ}[t][l] \tag{4.3}$$

where T is the exponential filter constant. This averaging aims at introducing the time steadiness of the PRB pattern. The instant quality metric on PRB l is defined by the arithmetic mean of the SINR estimate over the different UEs p in the dB domain:

$$\mathbf{IQ}[t][l] = \overline{\widehat{SINR}_{dB}[t][\cdot,l]}_p \tag{4.4}$$

Under certain assumptions described below, $\mathbf{IQ}[t][k]$ ranks the different PRBs k by magnitude of interference with a certain degree of precision that depends on N_{UE}.

cond 1: The eNode-B always transmits the same power per PRB when a PRB is in the 'transmission' state. The SINR per UE p and PRB l at TTI t can be decomposed in the following way:

$$SINR[t][p,l] = \frac{\mathbf{g}[t][p] \cdot \mathbf{h}[t][p,l]}{\mathbf{l}[t][p,l]} \tag{4.5}$$

where $\mathbf{g}[t][p]$ is the received signal power. According to *cond 1:*, $\mathbf{g}[t][p]$ does not depend on the PRB l. $\mathbf{l}[t][p,l]$ is the total interference and noise. $\mathbf{h}[t][p,l]$ is the signaling fast fading component:

$$\mathbf{h}[t][p,l] = \sum_{a=1}^{N_{Rx}} \mathbf{HH}^S[a][t][p,l] \tag{4.6}$$

Where S is the index of the signaling link. Then assuming a perfect SINR estimate, the instant quality metric can be decomposed as:

$$\mathbf{IQ}[t][l] = \overline{\mathbf{g}_{dB}[t][\cdot]}_p + \overline{\mathbf{h}_{dB}[t][\cdot,l]}_p - \overline{\mathbf{l}_{dB}[t][\cdot,l]}_p \tag{4.7}$$

cond 2: All UEs of the cell undergo the same fast fading conditions: $\forall$ *PRB* l, $\mathbf{h}_{dB}[t][p,l]$ *are Independent and Identically Distributed (i.i.d) among UEs* p. *Furthermore* $\mathbf{h}_{dB}^{max}$ *is the unique value that maximizes the probability density function of* $\mathbf{h}_{dB}[p,l]$. The uniqueness of the maximum of the probability density function of $\mathbf{h}_{dB}[p,l]$ is a reasonable assumption. Indeed, with a standard maximum ratio combining receiver, $\mathbf{h}[p,l]$ is chi-square distributed with a degree of freedom that depends on the number of receiving antennas. A chi-square distribution is either:

- strictly decreasing for 1 and 2 degrees of freedom and therefore in that case accept 0 as unique maximum,
- or admitting a maximum in $2 \cdot \left(\frac{k}{2} - 1\right)$ for $k > 2$ where k is the number of degrees of freedom.

With *cond 2*, we can apply the central limit theorem with two consequences. Firstly, $\overline{\mathbf{g}_{dB}[\cdot]_p} + \mathbf{h}_{dB}^{max} - \mathbf{IQ}[l]$ is a maximum likelihood estimate of the average interference on PRB k: $\overline{\mathbf{l}_{dB}[\cdot,l]_p}$. Secondly:

$$\lim_{N_{UE}\to+\infty} \text{std}\left(\overline{\mathbf{h}_{dB}[\cdot,l]_p}\right)_t = 0 \tag{4.8}$$

Therefore, the precision of the estimate increases with the number of UEs in the cell N_{UE}. Finally, as $\overline{\mathbf{g}_{dB}[\cdot]_p} + \mathbf{h}_{dB}^{max}$ does not depend on l, $\mathbf{IQ}[l]$ ranks the PRBs in ascending order from highest interference to lowest interference.

4.2.4 Reuse $\frac{1}{3}$

Reuse $\frac{1}{3}$ is an improvement to the three previously presented PRB pattern selection techniques aiming at improving the general SINR conditions of the network by prioritizing different zones of the bandwidth for transmission depending on the antenna orientation. In our layout model, as shown in figure 2.1, there are three types of sectors. Each type of sector has a specific type of antenna orientation and is surrounded only by sectors of the two other types. Figure 4.3 describes the reuse $\frac{1}{3}$ principle. It consists in dividing the transmission bandwidth in three adjacent zones of equal size. Each zone is associated with a sector type as being the scheduling priority zone. In our 50 PRBs Evolved Universal Terrestrial Radio Access Network (E-UTRAN) modeling, the prioritized zone consists of a set of PRB indices Γ defined as follows:

$$\Gamma = \begin{cases} [1,17] & \text{for sector of type s1} \\ [18,34] & \text{for sector of type s2} \\ [35,50] & \text{for sector of type s3} \end{cases} \tag{4.9}$$

If all sectors are transmitting exclusively over their priority zone, then no interference is created by adjacent sectors. This is illustrated in figure 4.3 where the sector of type S2 is surrounded only by sectors of types S1 and S3 which have scheduling priority zones orthogonal to the scheduling priority zone of sectors of type S1.

As BM or BQPP selects at TTI t the N_{PRB}^{Tx} PRBs that maximize a value $\mathbf{Q}_X[t][l]$ (where X stands for $BQPP$ or BM), Reuse $\frac{1}{3}$ can be easily combined with BM and BQPP as follows. We can apply the following transformation to $\mathbf{Q}_X$:

$$\mathbf{Q}_X^{R\frac{1}{3}}[t][l] = \begin{cases} f^{[1,2]}(\mathbf{Q}_X[t][l]) & \text{if } l \in \Gamma \\ f^{[0,1]}(\mathbf{Q}_X[t][l]) & \text{if } l \notin \Gamma \end{cases} \tag{4.10}$$

where $f^{[a,b]}$ is a function with the following properties:

- $f^{[a,b]}$ is defined over $\mathbb{R}$ and return values in $[a,b]$,
- $f^{[a,b]}$ is strictly increasing over $\mathbb{R}$.

The transformation separates the PRBs in two sets giving strictly higher $\mathbf{Q}_X^{R\frac{1}{3}}$ values to PRBs in Γ compared to PRBs that are not in Γ while keeping the same order than $\mathbf{Q}_X$ within each set. $\mathbf{Q}_X^{R\frac{1}{3}}$ therefore ranks the PRBs first according the the Reuse$\frac{1}{3}$ principle and then according the the BM

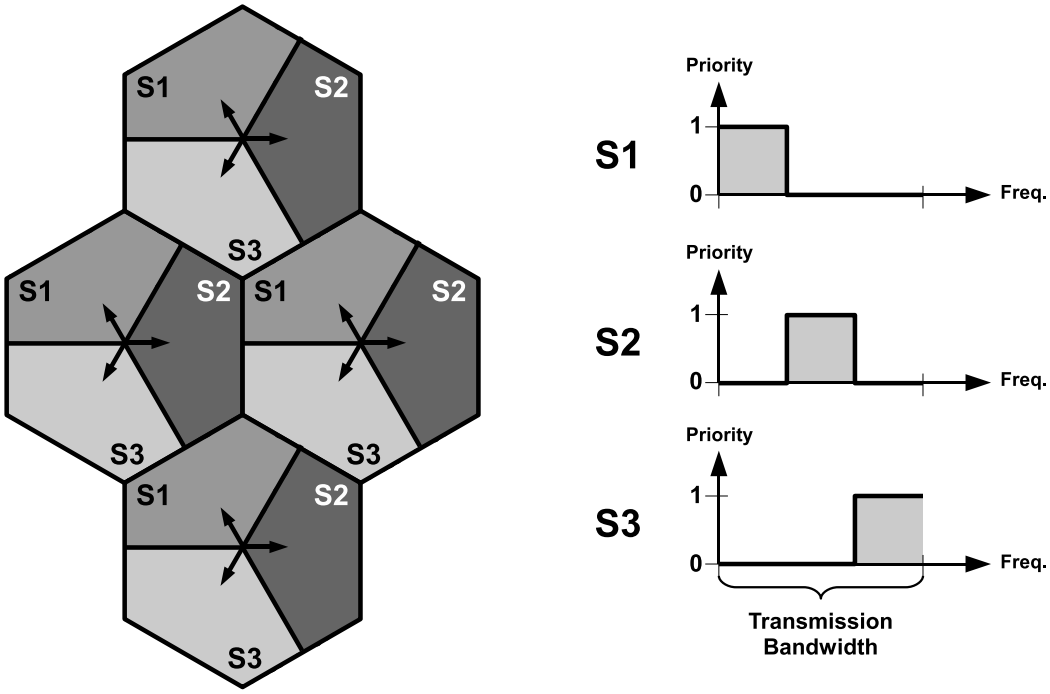

Figure 4.3: Reuse 3 prioritizing pattern

or BQPP algorithm. Finally, applying the Reuse$\frac{1}{3}$ algorithms combined with BM or BQPP will simply consist of choosing at TTI t the N_{PRB}^{Tx} PRBs that maximize a value $\mathbf{Q}_X^{R\frac{1}{3}}[t][l]$ (where X stands for $BQPP$ or BM).

4.2.5 Wideband Interference Reporting

We have mentioned in section 4.1 that a negative effect of a fractional load scenario could be a too fast variation of the interference leading to the non relevance of the CQI due to reporting delays. We have proposed as solution with the RCPP and BQPP algorithms to force a certain steadiness in time of the transmission PRB pattern. We propose here an alternative solution. Wideband Interference Reporting (WIR) has been proposed in 3rd Generation Partnership Project (3GPP) in order to overcome frequency wise reference signal measurement imprecisions due to the fast changing of Multiple Input Multiple Output (MIMO) precoding matrices. We believe that wideband interference reporting can also have an critical effect on fractional load scenarios.

WIR consists in reporting the frequency domain CQIs calculated with an average interference instead of the local interference. In our model, the ideal CQI would therefore be calculated as follows (instead of (2.6)):

$$\mathbf{CQI}_{WIR}^{id}[m] = \tag{4.11}$$

$$\sum_{a=1}^{N_{Rx}} \frac{\sum_{k=(m-1)\cdot N_{CQI}+1}^{m\cdot N_{CQI}} \mathbf{P}^S[k] \cdot L_{ant}^S \cdot L_{path}^S \cdot S^S \cdot \mathbf{HH}^S[a][k]}{\frac{N_{CQI}}{N_{PRB}} \sum_{k=1}^{N_{PRB}} \left(\sum_{i \neq S}^{N_{sec}} \left(\mathbf{P}^i[k] \cdot L_{ant}^i \cdot L_{path}^i \cdot S^i \cdot \mathbf{HH}^i[a][k] + W_f \cdot W_{th} \cdot Bdw_{sub} \right) \right)} \tag{4.12}$$

The main difference with (2.6) is in the denominator. Indeed, in (4.11), the interference is calculated as the average of all PRB's interference. WIR aims at insuring at any time a certain level of accuracy of the CQI report in a fractional load situation by averaging out the possible CQI errors due to delay. However, though wideband interference reporting may avoid critical CQI inaccuracies, it introduces an error due to the interference averaging. The magnitude of this error increases while the coherence bandwidth of the channel decreases. We expect these errors to be compensated by the Outer Loop Link Adaptation (OLLA).

Note that wideband interference reporting provides information about the overall magnitude of the interference but does not provide information about the frequency variations of the interference. Therefore, it would not make sense to use WIR combined with BQPP which is based on the utilization of the frequency interference variation information.

4.2.6 PRB pattern selection method evaluation

In order to evaluate the performance of the different PRB pattern selection methods, the set of simulations described in table 4.1 are run. Different values of BFF are simulated spanning from low load ($BFF = 0.25$) to full loading ($BFF = 1$). For all the simulations, we keep the ratio between the number of users and fraction of the bandwidth used constant:

$$\frac{N_{UE}}{BFF} = 20 \tag{4.13}$$

Hence, the bandwidth available per UE remains constant for all the simulations. It is reminded here that for the different BFF values indicate the portion of the bandwidth in use. This portion does not vary in time. Furthermore, the PS used is resource fair. Indeed, the Round Robin (RR)-Time Domain Packet Scheduling (TDPS) gives exactly an equal time share of the resource to each UE. The Carrier over Interference to Average (CoItA)-FDPS as seen in Chapter 3 distributes resources fairly while scheduling UEs with the highest fade. The constraint on the number of users expressed in (4.13) combined with the resource fair PS strategy provides a simulation framework where every UE in each simulation is provided approximately the same share of the bandwidth. This enables to compare fairly the influence of the PRB pattern selection over the cell throughput and the coverage.

Figure 4.4 is an overview of the simulation results in terms of coverage and spectral efficiency. The spectral efficiency SE is defined here by:

$$SE = \frac{\overline{R}_{cell}}{BFF \cdot BdW} \tag{4.14}$$

Table 4.1: Simulation parameters for the evaluation of the different PRB pattern selection algorithms

Parameter	**Setting**
Envirronment	MACRO #1
Layout Configuration	19 simulated site wrap around
Traffic models / call arrival	Finite Buffer T_{FB} = 80 / 40 / 27 / 20 N_{UE} = 20 / 15 / 10 / 05 B_{FB} = 2.0Mbits
TDPS	RR
FDPS	CoItA
Fractional Load scenario	constant BFF BFF = 1.00 / 0.75 / 0.50 / 0.25
PRB pattern selection algorithms	BM RCPP BQPP BM - Reuse$\frac{1}{3}$ BQPP - Reuse$\frac{1}{3}$ WIR - BM WIR - RCPP WIR - BM - Reuse$\frac{1}{3}$

The results for full bandwidth utilization ($BFF = 1$) are considered as the reference results. Of course, for $BFF = 1$, only the nature of the interference reporting (localized interference reporting or WIR) influences the results. Indeed, as all PRBs are systematically in transmission mode, the PRB pattern selection algorithms (BM, BQPP, RCPP or Reuse $\frac{1}{3}$) do not change the results. By observing the overall results, it is concluded that:

- For WIR simulation, the coverage and spectral efficiency don't change significantly depending on the PRB pattern selection method;
- Combining a PRB selection method with Reuse$\frac{1}{3}$ does not change significantly the coverage and spectral efficiency results.

Therefore, the overall results can be summarized by figure 4.5 which shows the gains of BM, RCPP, BQPP and WIR with different BFF values over the full bandwidth utilization case ($BFF = 1$) with localized interference reporting.

A clear hierarchy appears between the different PRB pattern selection methods. BM, the simplest algorithm leads to a dramatic drop in spectral efficiency and coverage while RCPP and BQPP bring a gain of up to approximately 100% in both coverage and spectral efficiency with $BFF = 0.25$. BQPP shows a gain slightly higher than RCPP. WIR, as RCPP and BQPP brings improvement in terms of coverage and spectral efficiency when BFF decreases, however, the gain is lower than RCPP and BQPP.

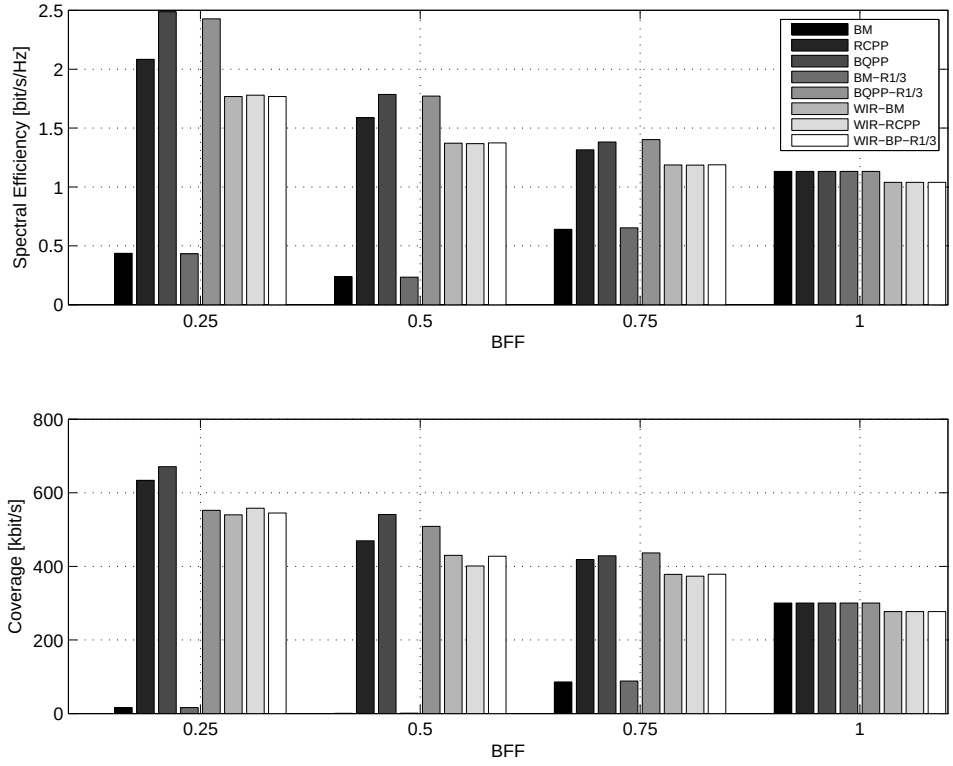

Figure 4.4: Coverage and Spectral efficiency with different PRB pattern selection algorithms.

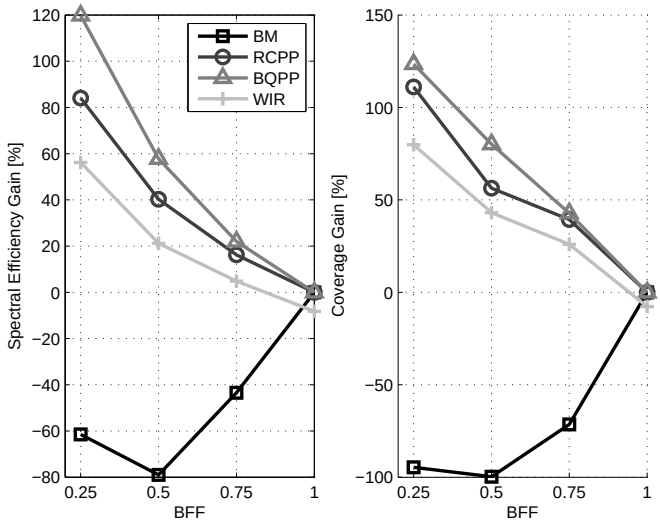

Figure 4.5: Coverage and Spectral efficiency with different PRB pattern selection algorithms.

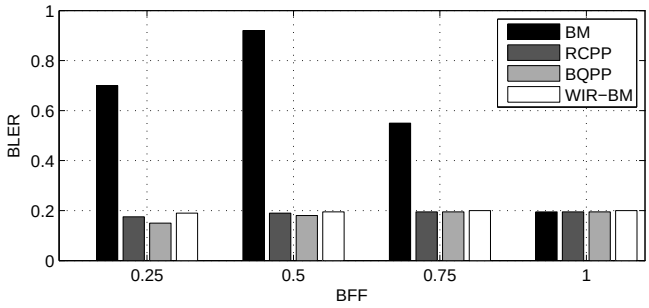

Figure 4.6: BLER with different PRB pattern selection algorithms for different values of BFF.

Figure 4.6 shows the different average first transmission BLER values for the different PRB pattern selection methods. BM is the only method that does not stick to the BLER target of 0.2 in a fractional load situation. BM reaches average BLER values systematically above 0.5 in fractional load scenarios. The highest average BLER values is 0.93 and occurs for $BFF = 0.5$. Such high values are not acceptable for the system to work properly as it means that Hybrid Automatic Repeat reQuest (HARQ) retransmissions are triggered almost systematically for each block transmission. The consequence are: increased transmission delays, and as shown on figure 4.4 low coverage and non efficient bandwidth usage. This high BLER is due to fast transitions of PRBs from 'transmission' to 'off' state and reversely, combined with the CQI transmission delay. Indeed as mentioned in Section 2.5.4, when delivered, CQI reports are several TTIs old. Focusing on one PRB:

- if the PRB was in 'off' state in neighboring cells during the TTI where the CQI report has been sent,
- but is in 'transmission' state in neighboring cells in the current TTI

then, the LA may overestimate the SINR of the PRB and therefore allocate a too high Modulation and Coding Scheme (MCS) to that PRB. The consequence is a highly probable block error. It is possible to illustrate and explain why fast PRB state transitions occur for BM by considering different interference situations in a group of N_{cells} neighboring cells. Those cells are geographically close, therefore they all interfer with each other. From a global system point of view, the 'total interference' created on a given PRB can be characterized by the number of cells in which this given PRB is in 'transmission' state. As in the cell layout used for the simulation described in Section 2.2 a significant number of cells (57 cells) are considered and those cells interfer with each other, the total interference and the interference from a cell point of view is therefore highly correlated. The BM method choses the PRBs with the best FDPS metric among all UEs. In the simulations run, the FDPS metric is CoItA, which according to the definition in Section 3.5.1 corresponds to the relative fade of the UE. However, considering the practical implementation of CoItA (3.21) used in a fractional loading situation, the CoItA metric is also scaled up by the SINR gain due to fractional loading. As BM used with CoItA will consider the best metric for each PRB, the states of the system can be summarized by the two following situations:

- *Situation 1:* In the N_{cells} cells, the 'total interference' is spread very unequally over the different PRBs.

In that situation, with BM combined with CoItA, all cells will tend toward selecting the PRBs with the least interference as the fractional load related SINR gain scaling will prevail in the CoItA

metric over the relative fade. Therefore, the 'total interference' will be again spread unevenly as all cells will react in the same way and transmit on the low interference PRBs thus creating high interference on those PRBs. This means that *Situation 1* will always tend to trigger *Situation 1* by systematically moving the zones of interference around the spectrum thus triggering fast PRB state transition.

- *Situation 2:* In the N_{cells} cells, the 'total interference' is spread evenly over the different PRBs.

In that situation, the ranking of the PRBs with the BM method will be based on the relative fade of the UEs with the best relative fade on each PRB. As the ranking is in that case not based on the interference and as the fading process is a random process, the PRB ranking will be chosen randomly regarding the interference. Therefore, when the system is in *Situation 2*, it is just a matter of time before it falls into *Situation 1*. Finally, the system should always converge toward *Situation 1* and therefore, tend toward a situation with fast changing PRB states and therefore with high BLER. As mentioned earlier, figure 4.6 shows that the BLER is the highest for $BFF = 0.5$. This can easily be explained by the fact that $BFF = 0.5$ is the situation that allows the highest number of PRB transition state every TTI: $\frac{N_{PRB}}{2}$.

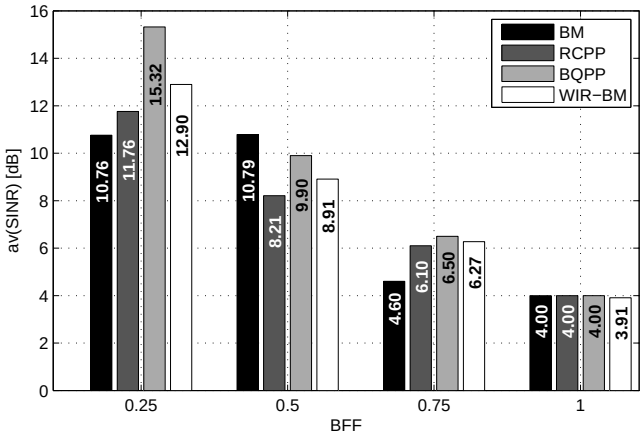

Figure 4.7: average SINR values with different PRB pattern selection algorithms for different values of BFF.

Fast PRB state transition is overcome with RCPP and BQPP by simply forcing PRBs to remain in the same state for a period of time significantly larger than the LA delay. Therefore, for RCPP and BQPP, the global SINR conditions improvement of the system when BFF decreases shown in figure 4.7 translates into coverage and spectral efficiency increase as shown in figure 4.5. Moreover, BQPP brings a significant SINR condition improvement compared to RCPP thanks to the PRB selection based on the lowest interference PRBs.

With WIR, the main difference with other schemes is that the eNode-B had no knowledge of the localized interference as the CQI report is based on wideband interference. The main consequence is that no PRB pattern selection based on the interference in different PRBs is relevant. Therefore, BM combined with CoItA does not trigger fast PRB state transition and BQPP cannot select the PRBs with the lowest interference. The selected PRB pattern is systematically chosen without any correlation with the interference pattern similarly to RCPP. From that point of view, any PRB pattern selection scheme is equivalent. Furthermore, the fast PRB state transition, which

are not triggered with WIR but that still can occur to a lesser extent with BM, don't have any influence on the system as the CQI is always reported with an average error. Therefore, the OLLA simply compensates by introducing a negative CQI offset which results in a global use of lower MCSs and a generally lower throughput and coverage compared to RCPP.

Finally, reuse$\frac{1}{3}$ has no influence of the different PRB pattern selection algorithms. This can be simply explained by the fact that reuse$\frac{1}{3}$ separates the PRBs only in two. The reasoning concerning BM still holds on the two parts of the spectrum and reuse$\frac{1}{3}$ cannot prevent fast PRB state transitions. A similar phenomenon holds for BQPP where the separation in two zones does not influence significantly the SINR improvement mechanism.

4.3 Coverage enhancement techniques

In Section 4.2, PRB pattern selection methods are studied in a context where their performance in terms of SINR and capacity gain can be highlighted. They have been proved to provide a consistent SINR gain when decreasing the BFF. In the present section, PRB pattern selection is studied as a mean to increase the system coverage by artificially reducing the proportion of the bandwidth in use. If reducing BFF will on one had provide SINR condition improvement, on the other hand, the system capacity will be reduced by the bandwidth limitation.

It has been shown in Section 3.7, that when the number of UEs is large compared to N_{MUX}, the TDPS Blind Equal Throughput (BET) scheduler combined with the FDPS metrics CoItA or Proportional Fair scheduled (PFsch) is a very throughput fair scheduler. Only little coverage improvement can be expected in that situation as in order to increase the coverage, the throughput of almost all UEs should be increased. However, when the number of UE is equal to N_{mux}, FDPS combined with weighting provides the best UE throughput fairness, but a large throughput difference remain between low G-factor UEs and high G-factor UEs as the range of throughput spans approximately over 1000kbps between 500kbps and 1500kbps. This is therefore in this configuration that coverage improvement by artificial bandwidth reduction can be expected. The simulation run are summarized in table 4.2. The effect of bandwidth reduction are tested with and without weighting in order to determine whether coverage gain could be cumulated. Only BQPP is tested here as it is the best PRB pattern selection method according to Section 4.2.6.

Figures 4.8 and 4.9 show the different throughput Cumulative Density Function (CDF) and the coverage and average cell throughput values. The results show that there is no coverage improvement by simply reducing the BFF. When no weighting is applied, the coverage is steady for $0.7 < BFF < 1$ while the average cell throughput drops of 10% between $BFF = 1$ and $BFF = 2$. When weighting is applied, the coverage decreases with BFF. A possibility to improve further the coverage thanks to a fractional load situation may be by designing a FDPS that has the ability to schedule low interference PRBs to low G-factor UEs in a better way than CoItA.

4.4 Bandwidth usage reduction for dynamic arrival scenario

In Section 4.3, fractional load is considered as a mean to improve coverage and therefore was intentionally created by the eNode-B. In the present Section, fractional load is considered as situation that occurs due to specific traffic conditions. The effect of fractional load on the system

Table 4.2: Simulation parameters for the evaluation of PRB pattern prioritizing as coverage enhancement method.

Parameter	**Setting**
Envirronment	MACRO #1
Layout Configuration	19 simulated site wrap around
Traffic models / call arrival	Finite Buffer T_{FB} = 20s N_{UE} = 10 B_{FB} = 1.0Mbits
TDPS	RR
FDPS	CoItA
Fractional Load scenario	CoItA with weight constant BFF BFF = 1.0 / 0.9 / 0.8 / 0.7
PRB pattern selection algorithm	BQPP

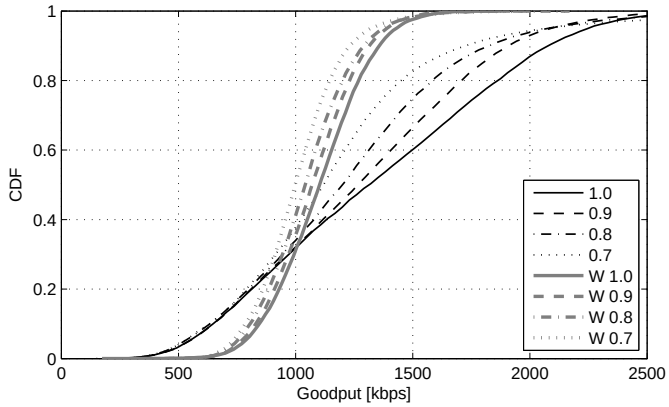

Figure 4.8: Coverage enhancement: throughput CDF with different BFF values. In the legend, W is mentioned for cases where weighting is applied

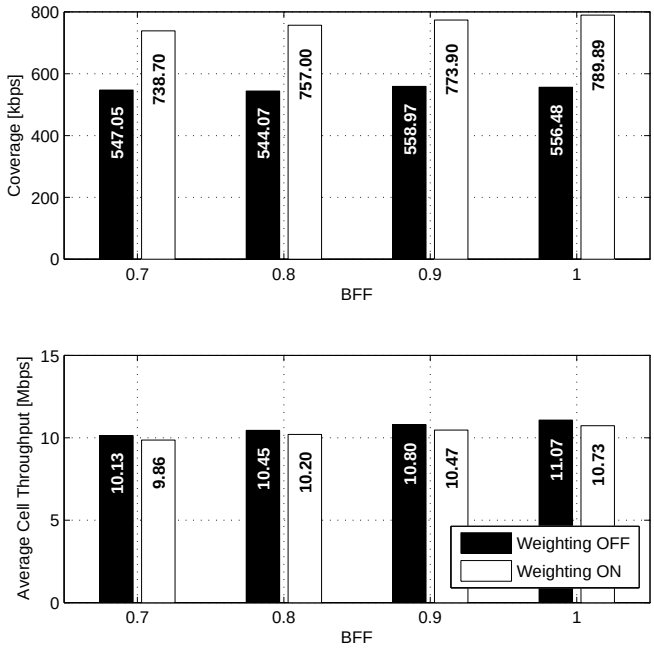

Figure 4.9: Coverage enhancement: Average cell throughput and coverage for different BFF values.

in various scenarios and the potential improvement brought by PRB pattern selection are studied.

4.4.1 Bandwidth use in low traffic conditions

In general, in cellular networks, cells are dimensioned and planned to accommodate the highest traffic peaks. However, situations where the capacity of the cell is higher than the overall traffic demand can occur. In those situations, there is theoretically no need to use the full bandwidth to provide to the UE all the offered traffic. A partial use of the bandwidth in case of low traffic load can help reducing the interference for neighboring cells thus increasing the capacity of those cells. In general, reducing the bandwidth use will improve the SINR conditions of the network. In order to clarify and define more precisely the notion of 'bandwidth use reduction', it is reminded that the PS algorithm is bound to the following constraints:

- *Constraint 1:* Every bit buffered at the eNode-B should be delivered if PRBs are available;
- *Constraint 2:* The chosen MCS aim at a BLER lower or equal to the BLER target.

Constraint 1 indicates that it is not intended to reduce the bandwidth when bits are available for transmission. *Constraint 2* is a built-in constraint of the LA algorithm. However, it is reminded here as it takes a special importance. Indeed, when facing a situation with a finite number of bits to be transmitted in a TTI, the minimum number of PRBs to use in order to transmit those bits is limited by this constraint. The matrix based PS algorithms described in Section 2.6.3 respects those two constraints as: when a UE has been scheduled on a set of PRBs that is sufficient to transmit all the bits of this UE with an MCS that complies with the BLER target this UE is removed from the list of schedulable UEs. If no more UEs are on the list of schedulable UEs, the algorithm

is terminated, therefore, the total bandwidth is not necessarily used. Finally, we can define the notion of bandwidth use reduction by the reduction that can be obtained by including PRB pattern selection in the matrix based PS algorithm, while respecting *Constraint 1* and *Constraint 2*.

4.4.2 From PRB pattern selection to PRB pattern prioritizing

Including PRB pattern selection in the matrix based PS algorithms differs from Section 4.2 in the following way:

- In Section 4.2, N_{PRB}^{Tx} is a parameter determined before the application of the PRB pattern selection algorithms and that helps determining a set of PRB for transmission;
- Instead, for bandwidth reduction as envisioned in Section 4.4.1, N_{PRB}^{Tx} is a consequence of the number of bits available for transmission in a TTI.

For this reason, the PRB pattern selection principle is softened here to PRB pattern prioritizing, which consists every TTI in creating i_{max} groups of PRBs:

$$\mathbf{S}^{PRB}[i]; \quad i \in [1, i_{max}]$$

where $\mathbf{S}^{PRB}[i]$ has a priority level decreasing when i increases. The matrix based algorithm is applied successively on the different PRB groups from the highest priority level to the lowest priority level. The algorithm terminates when no more UE has any bit left to transmit or when all PRBs are used. The algorithm is described in Figure 4.10.

4.4.3 Delay aware Packet Scheduler

In order to evaluate PRB pattern prioritization as a bandwidth usage reduction method, the Constant Bit Rate (CBR) traffic model is used as it can emulate situation with small buffers. If we assume that in low traffic situation, all UEs will be provided their CBR, this should results in a situation where:

$$\forall n \in [1, N_{UE}] \quad \overline{\mathbf{R}}[t][n] \approx \mathbf{CBR}[n] \tag{4.15}$$

In that case, the TDPS-Proportional Fair (PF) scheduler would be equivalent to a maximum throughput scheduler, which is very unfair resource-wise and therefore throughput-wise. Therefore it is preferred here to introduce a delay aware PS to handle the CBR traffic. The Modified Largest Weighted Delay First (M-LWDF) introduced in [41] and also studied in detail in [42] has proved to be a fair solution for handling delay sensitive traffic. The M-LWDF metric can be written as following:

$$\mathbf{M}_{M-LWDF}^{TD}[t][n] = \frac{\hat{\mathbf{D}}[t][n]}{\overline{\mathbf{R}}[t][n]} \cdot \Lambda[t][n] \tag{4.16}$$

Where $\Lambda[t][n]$ is the head of line delay of UE n at TTI t. The head of line delay is defined by the time during which the oldest packet stored in UE n's buffer has been waiting for complete transmission. Note that contrarily to the formulations in [41] and [42], (4.16) does not include the discard timer and priority weight components. Indeed, as the study does not aim at showing effect of discarding late packets, and as it includes only cases with single CBR and therefore no need or reasons for prioritizing certain UEs over other UEs, those two components would have no effect on the algorithm.

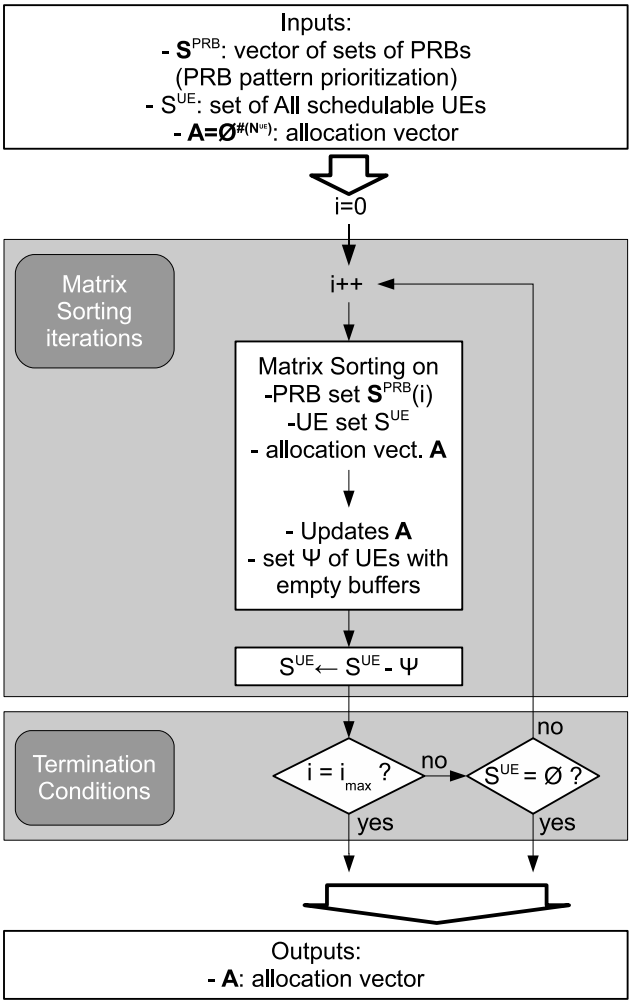

Figure 4.10: Description of the modified PRB allocation algorithms with integrated PRB prioritization

4.4.4 Simulation Results

In the evaluation of PRB pattern prioritization as a bandwidth usage reduction technique, two dimensions of the traffic are explored: the burstiness and the offered load. Firstly the burstiness is defined here by the frequency of data delivery at the eNode-B. For example a traffic that delivers large packets sparsely is considered as bursty while a traffic that delivers small packet very often is considered non bursty or smooth. This dimension of the traffic is critical regarding fractional load handling. Indeed, the number of PRBs that need to be in use depend on the number of available bits. The burstiness of the traffic will influence the number of bits available versus time and therefore, will influence the result of the bandwidth usage reduction methods. Then, the offered load is defined as the total throughput of the traffic incoming at the eNode-B. It is the traffic "offered" by the eNode-B for transmission to UEs. The offered load will influence directly the need for bandwidth. As seen in Section 4.2, the behavior of the PRB pattern selection methods changes with BFF.

Table 4.3 summarizes the different simulations run to test the bandwidth usage reduction methods. It has been chosen to test two different offered load: 3Mbps and 7Mbps. Those offered load are smaller than the cell capacity that is approximately 12Mbps as shown in section 2.6.5.2, there-

Table 4.3: Simulation parameters for the evaluation of PRB pattern prioritization as Bandwidth usage reduction method.

Parameter	**Setting**
Envirronment	MACRO #1
Layout Configuration	19 simulated site wrap around
Traffic model	CBR CBR = 512kbps T_{CBR} = 6s P_{CBR} = 256kb / 2.56kb / 256kb / 2.56kb
Call Arrival	Poisson AOL = 3Mbps 7Mbps $T_{Poi.}$ = 20s 10s
TDPS FDPS	M-LWDF CoItA
PRB pattern selection algorithms	BM BQPP BM-Reuse$\frac{1}{3}$ BQPP-Reuse$\frac{1}{3}$

fore, they should result in situations where the full bandwidth is not needed. The UE are generated with a CBR traffic of 512kbps but two different degrees of burstiness are tested. A bursty version with inter packet arrival time of 500ms (and packets of 256kbs) and a smoother version with inter packet arrival time of 5ms (and packets of 2.56kbs). Those four cases are tested with BM and BQPP and together with reuse$\frac{1}{3}$ as well.

Figures 4.11 and 4.12 summarize the simulation results. Figure 4.11 shows the bandwidth usage with different algorithms as the CDF of the number of PRBs used for transmission. Figure 4.12 shows the average BLER versus G-factor.

Firstly, observing the bursty UEs bandwidth usage on figure 4.11, it is concluded that those types of UEs operate mainly in two transmission configurations: 'full load' ($BFF = 1$) and 'zero load' ($BFF = 0$). Only during a small time proportion, the system is operating at an 'intermediate load' ($0 < BFF < 1$). Table 4.4 summarizes the percentage of time spent in those three different states. As the different PRB pattern prioritization methods have a significant impact on the system only in 'intermediate load', their effect is very limited in bursty traffic conditions. The results for bursty traffic used together with PRB prioritization methods are not shown on the different figure as their impact is indeed insignificant.

On the contrary, PRB prioritization methods have a significant impact on the non bursty traffic type as this traffic operates mainly in 'intermediate load'. Figure 4.11 shows clearly that for an offered load of 3Mbps, the different PRB prioritization methods enable a bandwidth usage reduction of 10 PRBs. For an offered load of 7Mbps, the bandwidth usage reduction is of only 3 PRBs on average. Figure 4.12 shows that the bandwidth reduction comes together with a global BLER decrease. The most obvious case comes for the offered load of 3Mbps where the BLER is

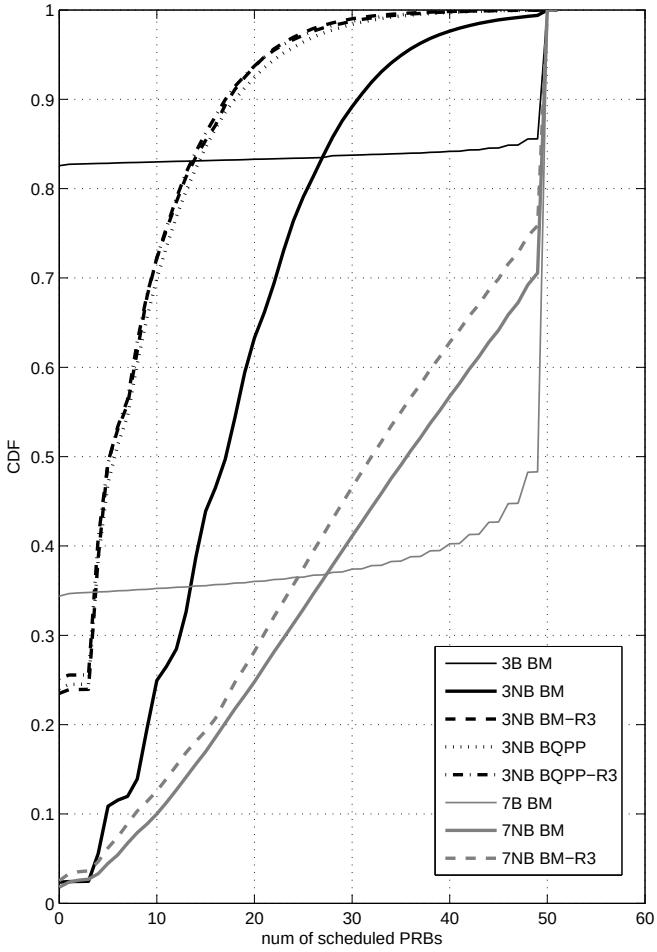

Figure 4.11: CDF of the number of PRBs used per TTI for different PRB pattern selection algorithms in different offered load and burstiness conditions. In the legend, NB stands for the non bursty configuration while B stands for the bursty configuration. The numbers 3 and 7 indicate the offered load in Mbps.

Table 4.4: Percentage of time spent in 'full load', 'zero load' and 'intermediate load' for the bursty traffic configurations.

Transmission Configuration	**3Mbps**	**7Mbps**
'Full Load'	14.5%	48.0%
'Zero Load'	82.5%	35.0%
'Intermediate Load'	3.0%	13.0%

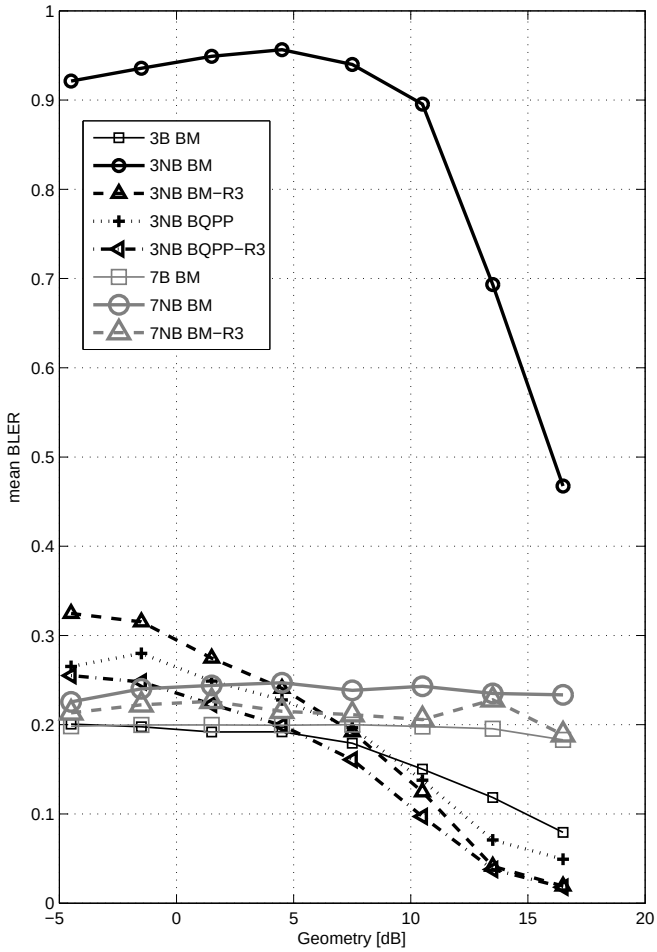

Figure 4.12: G factor versus average BLER for different PRB pattern selection algorithms in different offered load and burstiness conditions. In the legend, NB stands for the non bursty configuration while B stands for the bursty configuration. The numbers 3 and 7 indicate the offered load in Mbps.

near 100% for all UEs. The PRB prioritization methods allow to reduce the BLER significantly for the offered load of 3Mbps. However, the BLER remains above target for low G-factor UEs. This is due to the fact that in PRB pattern prioritization, BFF is not constant, as a consequence PRB state transitions occur depending on the bandwidth need and that effect cannot be controlled by PRB pattern prioritization. For an offered load of 7Mbps, the bandwidth reduction is less significant as the bandwidth usage is higher than for 3Mbps on average.

4.5 Conclusion

When fractional load occurs in a system due to low traffic offered load, it is shown that if not handled, the BLER can increase dramatically which can affect the Quality of Service (QoS) negatively by increasing packet delays. This situation happens especially if the type of traffic is very smooth in the sense that it is delivering relatively small packets compared to the average transport block size in one TTI.

In this Chapter, mainly two solutions to overcome the BLER increase are proposed. Firstly, PRB pattern selection methods trying to introduce correlation in the PRB usage prove to eliminate the BLER increase. The method Best Quality PRB Pattern (BQPP) stands out as it chooses the PRBs with the lowest interference and provides the best spectral efficiency gain (at 25% load, the spectral efficiency gain is of 120% compared to the full load case). Seconly, Wideband Interference Reporting (WIR) in the simplest solution and proves to also eliminate the BLER increase. Though WIR does not provide the same spectral efficiency gain than PRB pattern selection methods, it is the prefered solution for its simplicity and for the fact that additional spectral efficiency is not needed in low offered traffic load conditions.

Furthermore, fractional load is studied as a coverage improvement method. Though BQPP brings a significant SINR improvement, It is concluded that by forcing the system to use less bandwidth (forced fractional load) the combination between BQPP and the presented PS cannot bring any coverage improvement to the system. Therefore, the algorithms presented in the thesis in the simulation conditions, cannot be used as Inter-Cell Interference Coordination (ICIC) method.

Chapter 5

QoS Aware Packet Scheduling for Multiple Traffic Types

5.1 Introduction

In Chapter, 3 Packet Scheduler (PS) is studied as a mean to control fairness between different best effort users while trying to maximize the cell capacity. It is shown that the Time Domain Packet Scheduling (TDPS)-Priority Set Scheduler (PSS) scheduler enables to control the throughput while Frequency-Domain Packet Scheduling (FDPS)-Proportional Fair scheduled (PFsch) and FDPS-Carrier over Interference to Average (CoItA) provide the highest cell capacity. In Chapter 4, PS is studied under fractional load conditions, which can occur with limited offered load and real time traffic. It is concluded that the share of the spectrum in use must be steady in time for the system to work at BLock Error Rate (BLER) target. The present Chapter tackles the mixed traffic case where two types of User Equipment (UE)s are present simultaneously in the system:

- Best effort UEs which are not subject to any throughput nor Quality of Service (QoS) constraint,
- real time UEs which are subject to Guaranteed Bit Rate (GBR) and delay constraints as QoS parameters.

The problem of PS with QoS constraints is approached with the following targets in mind:

- *Target 1* With a given UEs configuration, the PS must be able to provide QoS to a maximum of UEs,
- *Target 2* The capacity of QoS UEs must be maximized.
- *Target 3* best-effort UEs don't have any specific QoS target. The system must share resource equally among best effort UEs.

Giving an equal share of the resource to all best effort UEs is not the strategy that maximizes the cell throughput. However, all best effort UEs must eventually be served, therefore, *Target 3* is a reasonable trade-off between coverage and cell capacity. Several PS strategies are introduced

and tested. Furthermore, the Required Activity Detection (RAD) concept is introduced as a mean to share the time domain resource combined with a weighted Proportional Fair (PF) algorithms but also as a mean to share the frequency domain resources combined with frequency domain weighting.

5.2 Generalization of QoS aware algorithms for multi traffic support

In this Section, the diverse algorithms presented in the previous chapters of the thesis are reformulated for multiple traffic with multiple QoS requirements support. Best effort UEs are set with a default GBR value: $GBR = 0$.

5.2.1 Normalized Blind Equal Throughput

The Blind Equal Throughput (BET) algorithm is designed to equalize the throughput among all UEs present in a cell. The Normalized Blind Equal Throughput (N-BET) algorithm aims at equalizing the UE throughput to GBR ratio. As best effort UEs have a GBR equal to 0, a minimum value GBR_{min}^{N-BET} is introduced in the nominator of the N-BET metric:

$$\mathbf{M}_{N-BET}^{TD}[t][n] = \frac{\max\left(GBR_n, GBR_{min}^{N-BET}\right)}{\overline{\mathbf{R}}[t][n]} \tag{5.1}$$

GBR_{min}^{N-BET} is a minimum target for best effort users. It must be set small enough so that the real-time UEs are given enough resource to comply with their GBR requirement. The setting of GBR_{min}^{N-BET} should therefore depend on the number of best efforts UEs present in the system and the requirements of the real-time UE. In order to illustrate the importance of the setting of GBR_{min}^{N-BET}, let us assume that a cell has a fixed capacity of C. If N_{BE} is the number of best effort UEs present in the system, then, GBR_{min}^{N-BET} should be subject to the following constraint:

$$GBR_{min}^{N-BET} < \frac{C - \sum_{n=1}^{N_{UE}} GBR_n}{N_{BE}} \tag{5.2}$$

As seen in Chapter 3, the capacity of the system depends on the scheduling strategy, therefore (5.2) cannot be applied strictly. However, it gives an understanding of the constraint linked to GBR_{min}^{N-BET}. Another property of N-BET is that it will try to equalize the throughput of best effort UEs, and therefore be in contradiction with *Target 3*.

5.2.2 Priority Set Scheduler

The PSS scheduler described in Section 3.4 is another alternative for performing QoS aware prioritizing for multiple traffics. TDPS-PSS is reformulated here to support best effort UEs:

$$\mathbf{M}_{PSS}^{TD}[t][n] = \begin{cases} f^{[1,2]}\left(GBR_n \cdot \frac{1}{\overline{\mathbf{R}}[t][n]}\right) & \text{if } \overline{\mathbf{R}}[t][n] < GBR_n \\ f^{[0,1]}\left(\max\left(GBR_n, GBR_{min}^{PSS}\right) \cdot \frac{\hat{\mathbf{D}}[t][n]}{\overline{\mathbf{R}}[t][n]}\right) & \text{if } \overline{\mathbf{R}}[t][n] \geq GBR_n \end{cases} \tag{5.3}$$

where $f^{[a,b]}$ described in Section 4.2.4, is a strictly increasing function of $\mathbb{R}$ and that returns values in $[a, b]$. TDPS-PSS has mainly two differences with TDPS-BET. Firstly, TDPS-PSS performs a strict prioritizing between best effort and real time UEs independently of the setting of any value. Indeed, if a UE does not comply with its GBR, it will fall into set 1 and be systematically prioritized over the UEs that comply with their GBR or best effort UEs. Moreover, the best effort UEs are scheduled according to the TDPS-PF metric and should therefore be given approximately equal shares of time. Besides, the value GBR_{min}^{PSS} in (5.3) determines the prioritizing policy between best effort UEs and real time UEs within set 2. It is chosen here to prioritize best effort UEs within set 2 by simply setting GBR_{min}^{PSS} to a high enough value so that the metric of bests effort UEs is always greater than that of the real time UEs.

5.2.3 Modified Largest Weighted Delay First

The Modified Largest Weighted Delay First (M-LWDF) TDPS has been introduced in Section 4.4.3 expressed in (4.16) as a delay aware TDPS. This formulation is not sufficient in order to support both real time and best effort users. In order to generalize the algorithms to the mixed traffic case, the utility framework for delay aware traffic introduced in [39] or similarly in [44] is shortly reminded. The delay aware PS can be formulated as follows:

$$\begin{aligned} \text{maximize } F\left(\mathbf{W}\right) &= \sum_{n=1}^{N_{UE}} U_n\left(\mathbf{W}\left[n\right]\right) \\ \text{subject to } \forall \quad n \quad \mathbf{W}\left[n\right] &> 0 \end{aligned} \tag{5.4}$$

where U_n is the utility function associated with UE n and $\mathbf{W}\left[n\right]$ the queuing delay of UE n. In order to provide the long term maximization in (5.4), the time domain packet scheduler must schedule the UEs maximizing the following metric:

$$\mathbf{M}_{Delay}^{TD}\left[n\right] = -U_n'\left(\mathbf{W}\left[n\right]\right) \cdot \frac{\hat{\mathbf{D}}\left[n\right]}{\overline{\mathbf{R}}\left[n\right]} \tag{5.5}$$

In the thesis, it is proposed to design two types of marginal utility functions for real time traffic and for best effort as follows:

$$U_{besteffort}'\left(\mathbf{W}\left[n\right]\right) = 0.5 \cdot 10^{-3} \tag{5.6}$$

$$U_{realtime}'\left(\mathbf{W}\left[n\right]\right) = 1 \cdot 10^{-3} + \mathbf{W}\left[n\right] \tag{5.7}$$

Finally, the time domain metric becomes:

$$\mathbf{M}_{Delay}^{TD}\left[t\right]\left[n\right] = \begin{cases} 0.5 \cdot 10^{-3} \cdot \frac{\hat{\mathbf{D}}[n]}{\overline{\mathbf{R}}[n]} & \text{if } GBR_n = 0 \\ 1 \cdot 10^{-3} + \Lambda\left[t\right]\left[n\right] \cdot \frac{\hat{\mathbf{D}}[n]}{\overline{\mathbf{R}}[n]} & \text{if } GBR_n > 0 \end{cases} \tag{5.8}$$

The metric described in (5.7) gives a nearly absolute priority to real time UEs. The priority of real time UEs increases as their head of line delay increases. Best effort UEs are served when not enough real time UEs are schedulable.

5.3 Time-Domain Required Activity Detection

5.3.1 Time Domain Resource Sharing

The time share $\eta[n]$ given to UE n is defined by the probability of being scheduled by the time domain scheduler *with a first transmission*. If the system is in a configuration where Link Adaptation (LA) and Outer Loop Link Adaptation (OLLA) are in a stable mode, a retransmission should occur for each UE with a probability of $BLER$. Therefore, the time share given to UE n can be formulated as:

$$\eta[n] = (1 - BLER) \cdot P(\delta[n] = 1) \tag{5.9}$$

where $\delta[n] = 1$ indicates that UE n is scheduled by the time domain scheduler and $\delta[t][n] = 0$ indicates on the contrary that UE n is not scheduled. The number of UEs scheduled by the TDPS is equal to:

$$N_{sch}^{TD} = \min(N_{sch}, N_{mux}) \tag{5.10}$$

where N_{sch} is the number of schedulable UEs. The total time share given to all UEs is determined by the number of scheduled UEs as follows:

$$\sum_{n=1}^{N_{UE}} \eta[n] = N_{sch}^{TD} \cdot (1 - BLER) \tag{5.11}$$

The maximum time share that the system can accommodate corresponds to the situation where N_{mux} UEs are scheduled every Transmission Time Interval (TTI). The total time share is therefore subject to the following maximum constraint:

$$\sum_{n=1}^{N_{UE}} \eta[n] < N_{mux}^{TD} \cdot (1 - BLER) \tag{5.12}$$

Additionally, the maximum time share of a UE is reached when it is scheduled every TTI, therefore, the time share is constrained as follows:

$$\forall n \quad \eta[n] < (1 - BLER) \tag{5.13}$$

Figure 5.1 illustrates the principle of time share with an example where $N_{sch}^{TD} = N_{mux}$. The time share of a UE is represented by a rectangle. The length of the rectangle corresponds its given time share while the width is constant and equal to 1, which corresponds to one UE. All the rectangles can be arranged in a bigger rectangle of length $BLER$ and width N_{mux}, which represents the total time share of the system.

5.3.2 Time Domain Weighted Round Robin

TDPS Weighted Round Robin is a very simple algorithm that aims at providing to each UE predefined target time shares $\eta^{Target}[n]$ of the resources to each UE n. It is defined by the following metric:

$$\mathbf{M}_{WRR}^{TD}[t][n] = \eta^{Target}[n] \cdot \tau[t][n] \tag{5.14}$$

where $\tau[t][n]$ is the number of TTIs since last schedule by the TDPS. If UE n is on average scheduled every time the metric $\mathbf{M}_{WRR}^{TD}[t][n]$ reaches 1, then UE n is provided time share $\eta^{Target}[n]$. This is possible of course only possible if the target time shares respect the conditions expressed in 5.12.

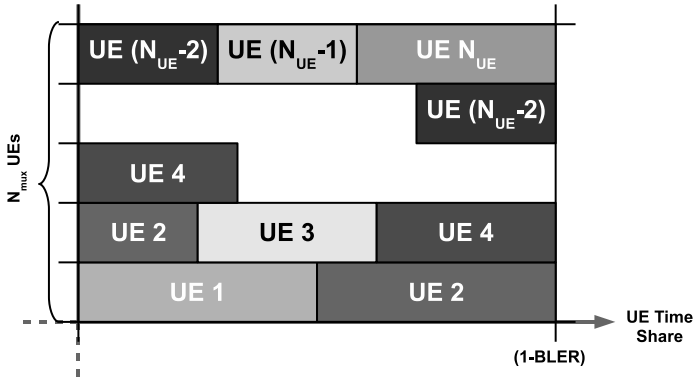

Figure 5.1: example of time sharing

5.3.3 Time-Domain Required Activity Detection

In the previous section, the concept of time resource share has been presented as well as a scheduler that aims at providing a desired time share to each UE. In that Section, the Time-Domain Required Activity Detection (TD-RAD) principle is completed by introducing a method to calculate the required time share of GBR UEs and the time share given to best effort UEs.

Firstly, it is assumed that the FDPS has a stationary behavior in the sense that the expected throughput given to any UE is constant. The expected throughput for UE n when scheduled by the TDPS is estimated by $\overline{\mathbf{R_{sch}}}\,[t]\,[n]$ defined in (3.23). Therefore, the time share required by GBR UEs can be estimated by:

$$\upsilon\,[t]\,[n] = \frac{GBR_n}{\overline{\mathbf{R_{sch}}}\,[t]\,[n]}. \tag{5.15}$$

By referring to the constraints specified in (5.12) and (5.13), satisfying all UEs with that time share is only possible if:

$$\sum_{n=1}^{N_{UE}} \frac{GBR_n}{\overline{\mathbf{R_{sch}}}\,[t]\,[n]} < N_{mux} \cdot (1 - BLER) \tag{5.16}$$

and:

$$\forall n \quad \upsilon\,[t]\,[n] < (1 - BLER) \tag{5.17}$$

If conditions (5.16) and (5.17) are satisfied, then all GBR UEs can potentially be provided their GBR. Furthermore, when the whole time share is then not fully attributed to GBR UEs, the excess time share is defined as follows:

$$\omega_{excess} = N_{mux} \cdot (1 - BLER) - \sum_{n=1}^{N_{UE}} \frac{GBR_n}{\overline{\mathbf{R_{sch}}}\,[t]\,[n]} \tag{5.18}$$

The excess time share can therefore be allocated to the $N_{UE/GBR=0}$ best effort UEs. In the TD-RAD concept, each best effort UE is given and equal share of the excess time domain resource. In order to formalize the sharing of the excess resource, a weight ω is defined:

$$\omega\,[t]\,[n] = \begin{cases} \frac{1}{N_{UE/GBR=0}} \left(N_{mux} \cdot (1 - BLER) - \sum_{n'}^{N_{UE}} \frac{GBR_{n'}}{\overline{\mathbf{R_{sch}}}[t][n']} \right) & \text{if } GBR_n = 0 \\ 0 & \text{if } GBR_n > 0 \end{cases} \tag{5.19}$$

Finally, the TDPS TD-RAD metric is defined as:

$$\mathbf{M}_{RAD}^{TD}[t][n] = (\upsilon[t][n] + \omega[t][n]) \cdot \tau[t][n] \tag{5.20}$$

Figure 5.2 illustrates the TD-RAD principle with $N_{mux} = 5$ and 4 best effort UEs.

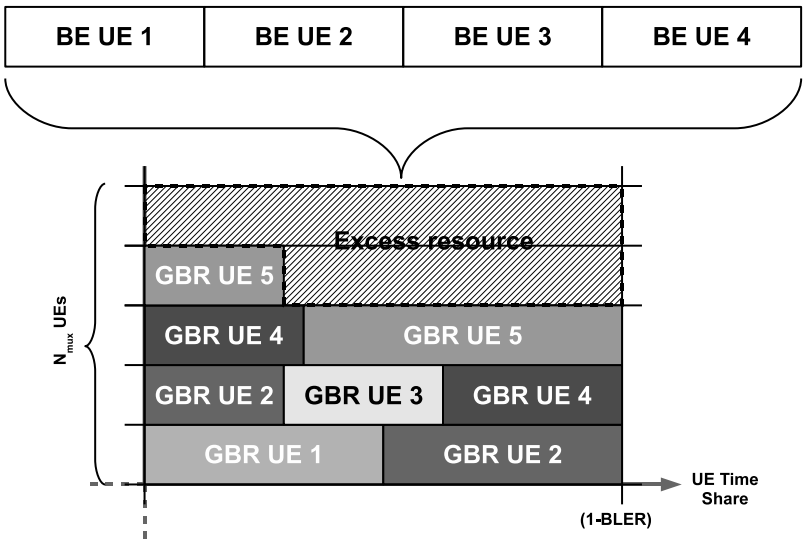

Figure 5.2: Example of time sharing with TD-RAD with $N_{mux} = 5$ and 5 GBR UEs and 4 BE UEs

5.4 Frequency-Domain Required Activity Detection

5.4.1 Frequency Domain Resource Sharing

In Orthogonal Frequency Division Multiple Access (OFDMA), the scheduling time share is not entirely relevant of the quantity of resource given to a UE. Indeed, as seen in Chapter 3, the average number of Physical Resource Block (PRB)s given to different UEs can vary depending on the G-factor. Therefore, the scheduling frequency share $\mu[n]$ of a UE n is defined as the average number of PRBs given to a UE when this UE is selected by the TDPS for a first transmission:

$$\mu[n] = E(\Delta[n] | \delta[n] = 1) \tag{5.21}$$

where $\Delta[n]$ is the number of PRBs allocated to UE n. The scheduling resource share of a UE can now be expressed as an average number of PRBs given over time by $\mu[n] \cdot \eta[n]$. With definitions (5.9) and (5.21) the total available resource share is therefore equal to:

$$\sum_{n=1}^{N_{UE}} \mu[n] \cdot \eta[n] = N_{PRB} \cdot (1 - BLER) \tag{5.22}$$

Furthermore, the maximum frequency share of a UE is reached when it is scheduled over the whole bandwidth. Therefore, the frequency share of a UE is bound as follows:

$$\forall n \quad \eta[n] < N_{PRB} \tag{5.23}$$

Finally, by combining with (5.13), the total scheduling resource share of a UE is constrained as follows:

$$\forall n \quad \mu[n] \cdot \eta[n] < N_{PRB} \cdot (1 - BLER) \tag{5.24}$$

Figure 5.3 illustrate the time and frequency scheduling share principle. each rectangle represents the scheduling resource share of a UE. the length represents the time share while the width represents the frequency share. The area of each rectangle corresponds therefore to the total scheduling share of the UE. All the rectangles can be arranged in a bigger rectangle that represents the total scheduling share of the system.

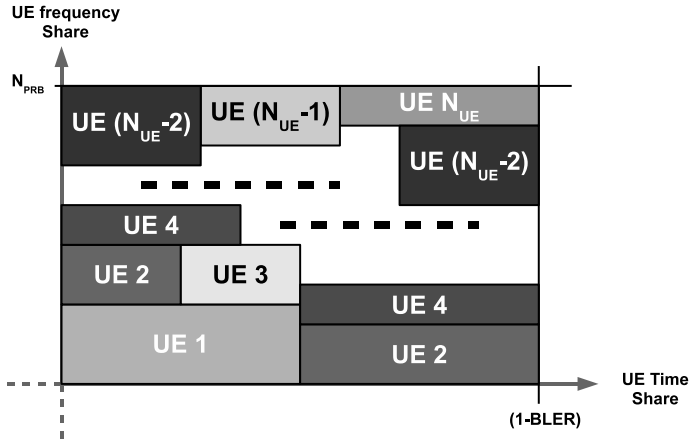

Figure 5.3: example of time sharing

5.4.2 Frequency Domain Metric Weighting

As seen in Chapter 3 a simple way to control the frequency share given to a UE is the frequency domain metric weighting of a specific UE. This simple method helps decreasing or increasing the average number of scheduled PRBs:

$$\mathbf{M}_{WX}^{FD}[t][n] = \mathbf{W}[n] \cdot \mathbf{M}_{X}^{FD}[t][n] \tag{5.25}$$

where $\mathbf{M}_{X}^{FD}[t][n]$ is X the FDPS metric and $\mathbf{W}[n]$ is the weight applied to it. It is not straightforward to design a type of weighting that will allocate a desired share of the resource. In the present thesis, the study is limited to a very simple weighting scheme where the UE weight consists of the frequency share intended for the UE.

5.4.3 Frequency-Domain Required Activity Detection

The TDPS-RAD algorithm described in section 5.3.3 aims at providing the exact time share required by GBR UEs so that they comply with their GBR requirement. However, TDPS-RAD cannot provide the GBR to a UE that is not given enough resource by the FDPS. Indeed, if

$$\overline{\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}}[t][n] < GBR_n \tag{5.26}$$

then, even if UE n is scheduled every TTI, it cannot be provided its GBR without increasing the frequency scheduling share given to UE n. FDPS-RAD aims at increasing the frequency

scheduling share to the UEs that don't fulfill condition 5.26 with the natural properties of the FDPS algorithms. The general principle consists in differentiating the UEs into two sets:

- *Set 1:*
  - The UEs that do not need to be scheduled every TTI and can comply with their GBR requirement,
  - the non GBR UEs,
- *Set 2:* the UEs that need to be scheduled every TTI and need to be allocated more PRBs compared to UEs of *Set 1*.

The algorithm proceeds as follows. Firstly, *Set 2* is composed of the UEs respecting the following criterion:

$$\overline{\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}}[t][n] < \alpha \cdot GBR_n \tag{5.27}$$

where α is slightly above 1. The value taken in the simulations is $\alpha = 1.1$. The reason for using α is that by applying FDPS-RAD, users in *Set 2* may reach $\mathbf{R_{sch}}[t][n] = GBR_n$ thanks to the weighted FDPS metric. However, in order to keep this situation where the UE is allocated enough resource to comply with its GBR, the UEs need to remain in *Set 2* and the weight needs to keep being applied. Therefore, *Set 2* consists of the UEs for which $\mathbf{R_{sch}}[t][n]$ is below the GBR or close to the GBR. The average number of PRBs required by UEs in *Set 2* in order to comply with their GBR requirement can be expressed as:

$$\boldsymbol{\Upsilon}[t][n] = \frac{GBR_n}{\overline{\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}}[t][n]} \tag{5.28}$$

where $\overline{\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}}$ describes the average throughput per PRB of UE n. $\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}[n]$ is calculated as follows:

$$\overline{\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}}[t][n] = \begin{cases} \frac{T_{sch}^{prb}-1}{T_{sch}^{prb}} \cdot \overline{\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}}[t-1][n] + \frac{1}{T_{sch}^{prb}} \cdot \frac{1}{\boldsymbol{\Delta}[t-1][n]} \cdot \mathbf{R}[t-1][n] & \text{if } \boldsymbol{\Delta}[t-1][n] \neq 0 \\ \overline{\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}}[t-1][n] & \text{if } \boldsymbol{\Delta}[t-1][n] = 0 \end{cases} \tag{5.29}$$

The total number of PRBs to be allocated by the FDPS to UEs in *Set 2* can be expressed by:

$$N_{W-PRB} = \sum_{n' \in Set2} \boldsymbol{\Upsilon}[t][n'] \tag{5.30}$$

Of course, similarly to TDPS-RAD with (5.16), the UEs in *Set 2* can be given the necessary resource only if

$$N_{W-PRB} < N_{PRB} \tag{5.31}$$

If condition (5.31) is respected, then all UEs in *Set 2* can be provided their GBR. In that case, N_{L-PRB} PRBs are left for UEs in *Set 1*:

$$N_{L-PRB} = N_{PRB} - N_{W-PRB} \tag{5.32}$$

In FDPS-RAD, it is chosen to share the number of PRBs evenly between UEs of *Set 1*. Therefore, a weight for each UE is defined as follows:

$$\boldsymbol{\Omega}[t][n] = \begin{cases} 0 & \text{if } n \in \textit{Set 2} \\ \frac{N_{L-PRB}}{N_{mux}-N_{Set\,2}} & \text{if } n \in \textit{Set 1} \end{cases} \tag{5.33}$$

where $N_{Set\ 2}$ is the number of UEs that fall into *Set 2*. The FDPS-RAD X metric is defined as:

$$\mathbf{M}^{FD}_{RAD-X}[t][n] = (\boldsymbol{\Upsilon}[t][n] + \boldsymbol{\Omega}[t][n]) \cdot \mathbf{M}^{FD}_{X}[t][n] \tag{5.34}$$

where X is the FDPS metric X. The FDPS-RAD principle is illustrated on figure 5.4 with 5 GBR UEs and 3 best effort UEs. In the figure, UE 1 and UE 2 are part of *Set 2*. Indeed, their time share has reached the maximum of $(1-BLER)$ and moreover, their frequency share has been increased so that they can be provided their GBR. Other GBR UEs and best effort UEs are part of *Set 1*. UEs in *Set 1* are not scheduled every TTI and share the rest of the resource. The maximum number of UEs scheduled per TTI remains constant but the number of PRBs allocated per UEs in *Set 1* as the number of PRBs allocated to UEs in *Set 2* increases.

Remarks:

- The algorithm aims at first detecting UEs that require extra PRBs (UEs in *Set 2*) and stabilize them within *Set 2* in a situation where they are provided enough PRBs.
- As the number of PRBs allocated to UEs of *Set 2* increases, the number of PRBs allocated to UEs in *Set 1* decreases. It is therefore possible that a UE that would originally be in *Set 1* fall into *Set 2* as a new *Set 2* UE comes in.
- FDPS-RAD relies on the fact that the FDPS weighted metric $\mathbf{M}^{FD}_{WX}[n]$ defined in (5.25) provides a frequency scheduling share to the UE n that is proportional to the weight $\mathbf{W}[n]$:

$$\mu[n] = N_{PRB} \cdot \frac{\mathbf{W}[n]}{\sum_{n'=1}^{N_{UE}} \mathbf{W}[n']} \tag{5.35}$$

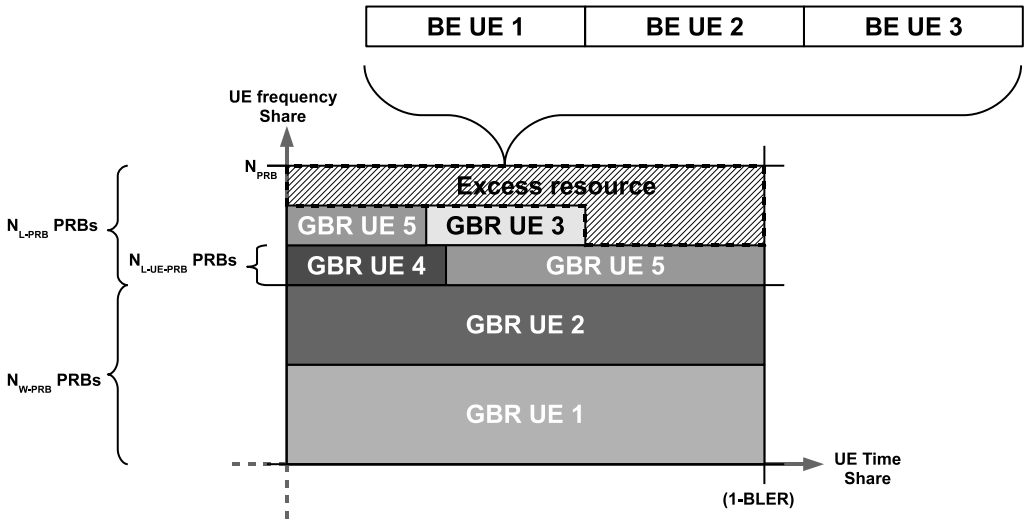

Figure 5.4: Example of frequency sharing with TD-RAD and FD-RAD. $N_{mux} = 5$ and 5 GBR UEs and 3 BE UEs

5.5 Performance Evaluation

In this section, the different algorithms and QoS handling concepts are evaluated as follows. Firstly, the focus is kept on the TDPS algorithms where the different algorithms are tested un-

der different offered loads. Then the effect of FDPS-RAD on scenarios with highly resource demanding UEs is evaluated.

5.5.1 Time Domain Scheduling QoS Differentiation

5.5.1.1 Simulations Cases

Table 5.1: Common simulation parameters for the evaluation of diverse QoS aware TDPS algorithms in mixed traffic conditions.

Parameter	**Setting**	
Envirronment	MACRO #1	
Layout Configuration	1 simulated site	
call arrival	N_{UE} constant	
	T_{FB} =	250s
Traffic Models	Finite Buffer	
	B_{FB} =	2.0Mbits
	GBR_{FB} =	0kbps
	Constant Bit Rate	
	CBR =	256kbps
	T_{CBR} =	10s
	P_{CBR} =	25.6kb
	GBR_{CBR} =	256kbps
TDPS	Priority Set Scheduler	
	Required Activity Detection	
	Blind Equal Throughput	
	Modified Largest Weighted Delay First	
	Proportional Fair	
FDPS	Carrier over Interference to Average	
	Proportional Fair scheduled	
	Proportional Fair	

Table 5.2: Traffic mixes cases for the evaluation of diverse QoS aware TDPS algorithms.

Parameter	**Setting**						
Cases	1	2	3	4	5	6	7
	$\gamma_{CBR} < 10$Mbps				$\gamma_{CBR} > 10$Mbps		
N_{UE} [UEs]	40				44	46	48
N_{UE}^{FB} [UEs]	40	27	14	1			
N_{UE}^{CBR} [UEs]	0	13	26	39	43	45	47
γ_{CBR} [Mbps]	0	3.33	6.66	9.98	11.00	11.52	12.03

Table 5.1 summarizes the common simulation parameters used for evaluating different QoS aware TDPS algorithms. The simulations run are of type "constant User Diversity Order (UDO) with mix of traffic" described in Section 2.3.3.1. The principle is reminded shortly here. Two types of traffics are used: finite buffer and Constant Bit Rate (CBR). Finite buffer UEs are considered as best effort UEs and their GBR is therefore set to 0. CBR UE emulate real time UEs and their GBR is therefore equal to their CBR. The finite buffer UEs have a buffer of 2Mbits and the CBR UEs have a CBR of $256kbps$ with a constant packet size of $25.6kbits$. in MACRO 1, $256kbps$ is low enough so that no UE undergo condition (5.26) and therefore, there is no need to apply FD-RAD nor any FDPS metric weighting (such as described in Section 3.7.3 or Section 5.4) in this context. This allows to focus only on TDPS here. The general call arrival scheme consists in keeping the total number of UEs within each traffic type (N_{UE}^{CBR} and N_{UE}^{FB}) constant thus keeping the total number of UEs N_{UE} constant as well. When a UE of one traffic type finishes its session, it is replaced by another UEs with same traffic type, placed randomly in the sector. With that type of call, the generated CBR traffic is equal to $\gamma_{CBR} = N_{UE}^{CBR} * CBR$ and γ_{CBR} is therefore called the CBR traffic offered load. On the contrary, the generated finite buffer traffic is not upper bounded as the UE buffer is available at the E-UTRAN Node B (eNode-B) as soon as the UE is in the network.

Table 5.2 summarizes the different traffic mix settings run. The configurations are designed keeping in mind that the cell capacity in MACRO 1 is around 10Mbps as shown in baseline results in Chapter 2. Firstly, 4 traffic mix settings consist of a CBR offered traffic load (γ_{CBR}) lower than 10Mbps and the total number of UEs is kept equal to 40. The goal of those simulations is to study how the different algorithms share the resource between finite buffer UEs and CBR UEs in a situation where the CBR traffic load can theoretically be supported by the sector. Then, 3 other simulation settings consist of a CBR offered traffic load greater than 10Mbps. For those settings, the number of finite buffer UEs is kept equal to 1 as the CBR load cannot theoretically be supported by a sector and therefore, only very little resource can be dedicated to finite buffer UEs. The goal of those settings is to study the behavior of the different TDPS algorithms when the CBR offered load is larger than the capacity.

Finally, the reference algorithm is the well-known TDPS-PF with FDPS-PF described in Section 2.6.3. The other TDPS algorithms tested are those presented in the present Chapter: BET, M-LWDF, PSS and TD-RAD. Those 4 TDPS algorithms are tested with the two FDPS algorithms presented in Section 3.7.2: PFsch and CoItA.

5.5.1.2 General Results

Figure 5.5 illustrates the effect of QoS aware PS: It shows the average cell throughput share of the two different traffic types in the different traffic configurations for two PS algorithms: TDPS-PF / FDPS-PF as a reference and TDPS-RAD / FDPS-PFsch. Firstly, it shows that for PF/PF, the cell throughput served to CBR traffic is lower than the offered CBR traffic. PF/PF is therefore unable to provide QoS to all UEs. On the contrary, the cell throughput served to CBR traffic with TD-RAD/PFsch corresponds exactly to the offered CBR traffic. In any case, none of the algorithms seems to be able to serve all the CBR traffic beyond an offered CBR traffic load of 10.5Mbps. Naturally, the throughput served to finite buffer UEs decreases when the the offered CBR load increases as the total resource is limited. The finite buffer traffic tends to fill up the excess resource left by the CBR traffic. Cell throughput representation of the results helps understanding the interaction between the different traffics and the role of the PS algorithms in terms of QoS provision. However, the real performance of the algorithms can only be evaluated in terms of

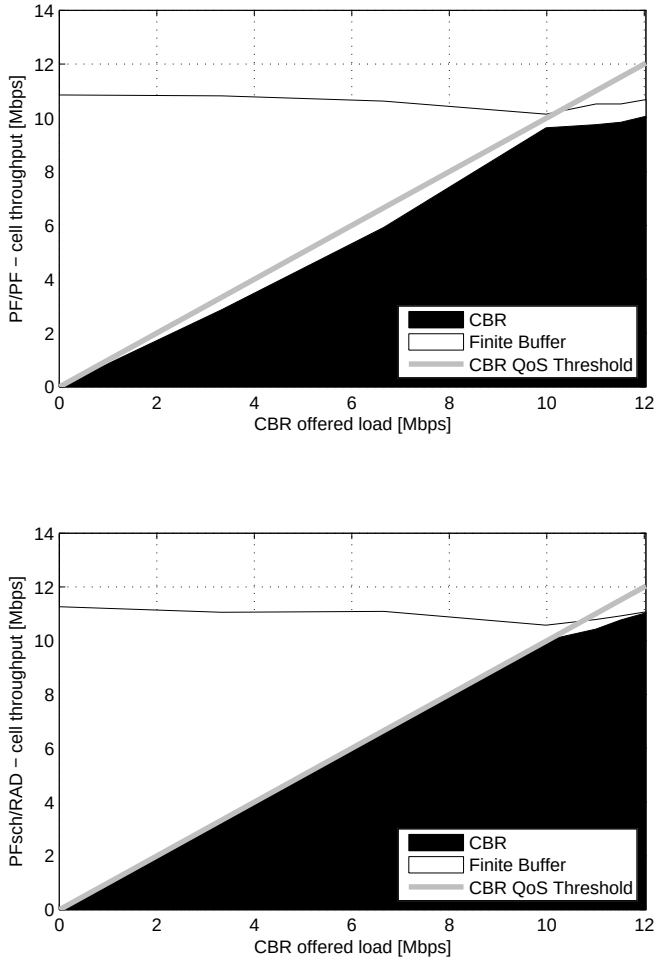

Figure 5.5: Examples of average cell throughput share between finite buffer and CBR traffic vs different CBR traffic offered loads for two PS algorithms: PF/PF and TD-RAD/PFsch.

QoS outage.

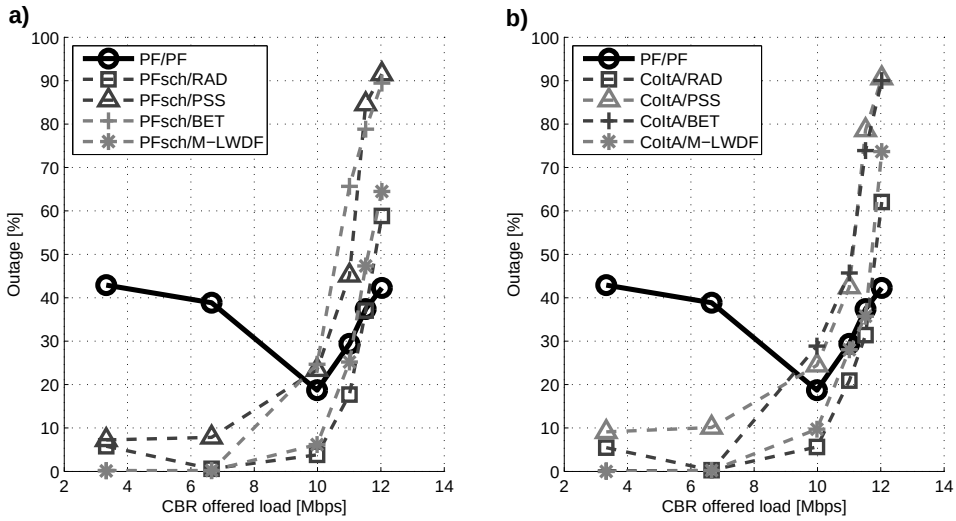

Figure 5.6: Outage values of the different TDPS algorithms. The outage is defined by the percentage of CBR UEs with a throughput inferior to their GBR. The graph on the left hand-side (a) provides results with FDPS-PFsch while the graph on the right hand side (b) provides results with FDPS-CoItA

Figure 5.6 represents the different GBR outage values of the different algorithms versus different CBR offered load. Results with FDPS-PFsch and with FDPS-CoItA are presented on two different graphs. The GBR outage is defined as the proportion of UEs that are not provided their GBR. More precisely, if the throughput delivered to UE n after the end of the session is $\mathcal{R}[n]$, UE n is in outage if:

$$\mathcal{R}[n] < (1-\beta) \cdot GBR_n \tag{5.36}$$

where β is close to 0. The purpose of introducing β is to relax the outage condition. Indeed, with the CBR traffic, the maximum achievable throughput is precisely equal to the GBR, therefore, β represents the small allowed throughput range. In the results presented in Figure 5.6, the value taken is:

$$\beta = \frac{255}{256} \tag{5.37}$$

Therefore, the UE that are not in outage are the UEs with a final throughput that ranges between 255kbps and 256kbps.

As apposed to cell throughput results, the outage represents the real performance of the algorithm in terms of QoS provision. Indeed, the cell throughput dedicated to CBR traffic does not give any information about the proportion of satisfied UEs as it does not indicate how this throughput is shared among UEs. Figure 5.6 shows three different tendencies in algorithms performance. Firstly, the PS/PF reference algorithms shows an outage that varies between 20% and 40% depending on the CBR traffic offered load. The minimum outage for PF/PF is met for $\gamma_{CBR} = 10Mbps$. As shown in many studies, the PF/PF algorithms is not suitable for mixed traffic as is is not able to keep the outage at a low value. Generally, the other algorithms have a relatively low outage ($< 10\%$) for $\gamma_{CBR} < 10$Mbps and the outage starts increasing very fast when γ_{CBR} increases above 10Mbps. However, for $\gamma_{CBR} > 10Mbps$, M-LWDF and TD-RAD show an outage lower by around 20 points compared to PSS and BET. Note that there is no change in tendency whether FDPS-PFsch or FDPS-CoItA is used.

5.5.1.3 The PF algorithm

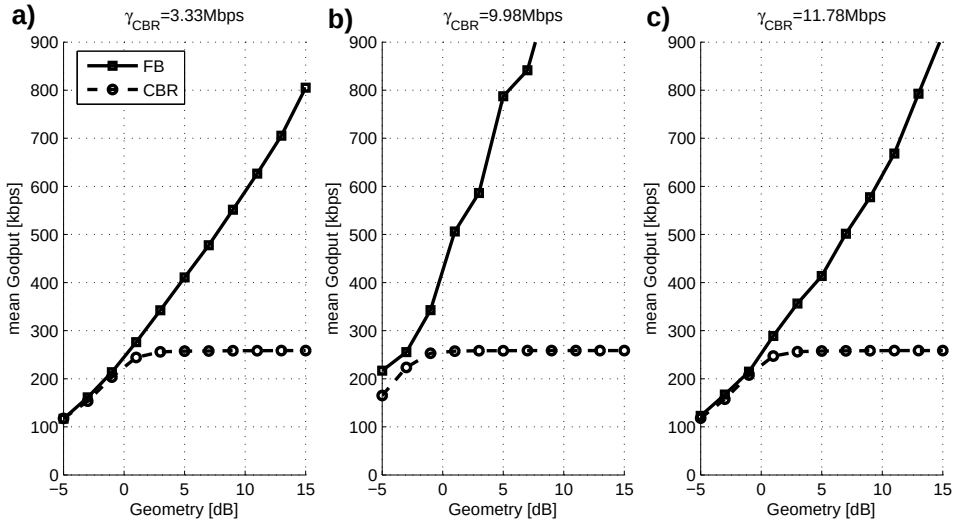

Figure 5.7: G-factor versus average Throughput for CBR traffic and for finite buffer traffic. The graph (a) shows results for $\gamma = 3.33Mbps$, (b) shows results for $\gamma = 9.98Mbps$, (c) shows results for $\gamma = 11.78Mbps$

Figure 5.7 shows the G-factor versus average throughput for CBR and finite buffer traffics for different CBR offered load values. From Figure 5.7, it is possible to highlight two types of behaviors regarding the TDPS-PF metric. Firstly, the UEs that are not limited by the their offered throughput belonging to a set that will be called S_{PF}^{-}. This includes all the finite buffer UEs as this traffic is not source limited by nature as well as all the CBR UEs for which the served throughput is inferior to the CBR ($\mathcal{R}[n] < GBR_n$). Figure 5.7 shows that there is a G-factor threshold $G_{\gamma_{CBR}}$ that depends on γ_{CBR} below which $\mathcal{R}[n] < GBR_n$. When the system is in a fixed state (with a fixed set of UEs in the cell), the TDPS-PF metric of the UEs of S_{PF}^{-} should converge toward a fixed value K_{PF}:

$$\forall n \in S_{PF}^{-} \quad \frac{\hat{\mathbf{D}}[n]}{\overline{\mathbf{R}}[n]} = K_{PF} \tag{5.38}$$

UEs of S_{PF}^{-} follow the PF principle and should therefore be allocated equal resource shares $RS_{s_{PF}^{-}}$:

$$\forall n \in S_{PF}^{-} \quad \mu[n] \cdot \eta[n] = RS_{s_{PF}^{-}} \tag{5.39}$$

With the power delay profile Typical Urban, which is used in all the simulations, the coherence bandwidth is small compared to the transmission bandwidth. The consequence is that the wide-band estimated throughput is approximately constant in time. Therefore, it is possible to consider $\hat{\mathbf{D}}[n]$ as nearly constant in time. Furthermore, $\hat{\mathbf{D}}[n]$ is an increasing function of the G-factor. The consequence is that $\overline{\mathbf{R}}[n]$ which according to (5.38) is proportional to $\hat{\mathbf{D}}[n]$, is also an increasing function of the G-factor. Finally, CBR UEs of S_{PF}^{-} with a G-factor below $G_{\gamma_{CBR}}$ are simply not provided their GBR. The total resource share allocated to the $N_{UE}^{S_{PF}^{-}}$ UEs of S_{PF}^{-}, $N_{UE}^{S_{PF}^{-}} \cdot RS_{s_{PF}^{-}}$ depends on the UEs with the second type of behavior regarding the TDPS-PF metric. They are the CBR UEs that are limited by their offered throughput (thus complying with their GBR requirements). Those UEs form a set that is called S_{PF}^{+}. As UEs of S_{PF}^{+} are throughput limited, their

past average throughput is equal to their GBR and therefore:

$$\forall n \in S_{PF}^{+} \quad \frac{\hat{\mathbf{D}}[n]}{\overline{\mathbf{R}}[n]} = \frac{\hat{\mathbf{D}}[n]}{GBR_n} > K_{PF} \tag{5.40}$$

The consequences are firstly that UEs of S_{PF}^{+} are systematically given priority over users of S_{PF}^{-} when they have data available at the eNode-B. Secondly, they are scheduled according to a max C/I rule. Paradoxically, S_{PF}^{+} UEs are given a smaller resource share than S_{PF}^{-} UEs since they are simply limited by their offered throughput.

Finally, the outage provided by the TDPS-PF algorithm depends on the balance between the two sets S_{PF}^{+} and S_{PF}^{-} where only UEs of S_{PF}^{+} are not in outage. Figure 5.6 shows that the outage provided by TDPS-PF decreases when γ_{CBR} increases for $\gamma_{CBR} < 10Mbps$ and increases for $\gamma_{CBR} < 10Mbps$. This trend can be explained by using a simple modeling of the resource sharing between S_{PF}^{+} and S_{PF}^{-} UEs. The total resource of the cell is shared as follows:

$$N_{UE}^{S_{PF}^{-}} \cdot RS_{s_{PF}^{-}} + \sum_{n \in S_{PF}^{+}} \mu[n] \cdot \eta[n] = 1 - BLER. \tag{5.41}$$

As UEs in S_{PF}^{+} are given a smaller resource share than S_{PF}^{-}:

$$\exists \quad \mathbf{p} \quad \text{so that} \quad \forall n \in S_{PF}^{+}, \quad 0 < \mathbf{p}[n] < 1 \quad \text{and} \quad \mu[n] \cdot \eta[n] + \mathbf{p}[n] = RS_{s_{PF}^{-}} \tag{5.42}$$

Therefore:

$$RS_{s_{PF}^{-}} = \frac{1 - BLER + \sum_{n \in S_{PF}^{+}} \mathbf{p}[n]}{N_{UE}} \tag{5.43}$$

For the simulation settings where $\gamma_{CBR} < 10Mbps$ (cases 1, 2, 3 and 4 described in Table 5.2), the total number of UEs is kept constant as mentioned in Table 5.1. In these conditions, increasing the proportion of CBR UEs by a factor of x should increase the number of UEs that fall into S_{PF}^{+} by a factor of at least x. Indeed, UE above the G-factor threshold should still fall into S_{PF}^{+} since they are traffic limited. (5.43) shows that on average, when the number of UEs in S_{PF}^{+} increases, then $RS_{s_{PF}^{-}}$ increases as well. If $RS_{s_{PF}^{-}}$ increases, this also means that the G-factor threshold decreases and that extra UEs will fall into S_{PF}^{+}. The result is the decrease in outage observed in figure 5.6 for $\gamma_{CBR} < 10Mbps$. On the contrary, for the simulation settings where $\gamma_{CBR} > 10Mbps$ (cases 5, 6 and 7 described in Table 5.2), N_{UE} increases together with the number of CBR UEs simultaneously in the network as mentioned in Table 5.1. In that case, $RS_{s_{PF}^{-}}$ decreases and the inverse effect regarding the outage is observed.

5.5.1.4 Feasibility Zone

The feasibility zone FZ_X^{δ} of a PS algorithms X is defined by the set of UE spatial configurations for which the outage is inferior to the value δ. In a wireless system, the Radio Admission Control (RAC) functionality is supposed to keep the system in the feasibility zone. The RAC is not the focus of this Chapter and is therefore not introduced before Chapter 6. Nevertheless, it is important to know the behavior of the different algorithms within their feasibility zone and outside their feasibility zone. Indeed, even if the RAC may prevent a new UE to make the system come outside of the feasibility zone, in a real system, UEs may be changing location and the system may come out of feasibility zone without any new UE entering the network. Determining precisely FZ_X^{δ} would require an infinite number of simulations, which is of course not possible within a PhD

study time. Therefore, in the simulation strategy, all CBR UEs have the same CBR and the focus is put on the value γ_X^δ, defined as the maximum average CBR offered load that can keep the outage inferior to δ for the PS algorithms X. The set of simulations run described in Table 5.1 does not allow to determine precisely γ_X^δ for any value of δ, however, it provides an interval where it is located. Considering $\delta = 0.1$, there is:

$$\gamma_{PSS}^{0.1}, \quad \gamma_{BET}^{0.1} \in [6.66Mbps, 9.98Mbps] \tag{5.44}$$

$$\gamma_{RAD}^{0.1}, \quad \gamma_{M-LWDF}^{0.1} \in [9.98Mbps, 10.75Mbps] \tag{5.45}$$

Hence the case with $\gamma_{CBR} = 9.98$ is special since it reveals the capacity gain brought by RAD and M-LWDF over BET and PSS.

5.5.1.5 Feasible Zone

In the present Section, the behavior of the different algorithms in the feasible zone is analyzed. this includes the simulation cases where $\gamma = 3.33Mbps$ / $N_{UE}^{BF} = 27$ and $\gamma = 6.66Mbps$ and $N_{UE}^{BF} = 14$. Figure 5.8 shows the G-factor versus average maximum delay for CBR traffic. The maximum delay of a UE is defined as the highest delay with which a packet is delivered to the UE. This measure helps understanding the behavior of the different TDPS algorithms with the CBR traffic. Furthermore, Figure 5.9 shows the Cumulative Density Function (CDF) of the number of schedulable UEs per TTI.

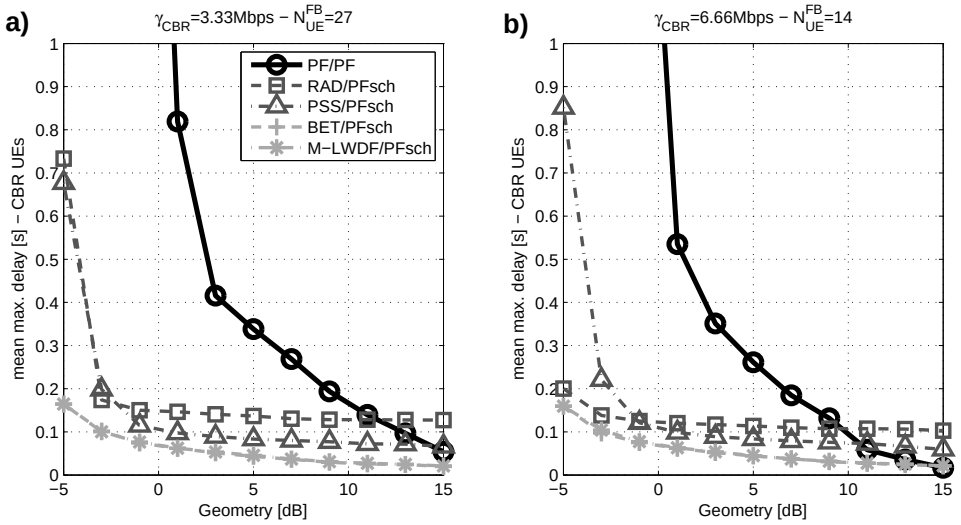

Figure 5.8: G-factor versus average maximum delay for CBR traffic. The graph (a) shows results for $\gamma = 3.33Mbps$ and $N_{UE}^{BF} = 27$, (b) shows results for $\gamma = 6.66Mbps$ and $N_{UE}^{BF} = 14$

BET and M-LWDF according to Figure 5.9, keep the number of schedulable CBR UEs below the number of UE that can be multiplexed in a TTI $N_{mux} = 10$. Note that by design, BET and M-LWDF give a nearly absolute priority to CBR UEs over best effort UEs. Therefore, in those conditions, as soon as a CBR UE is schedulable, it is scheduled. In the feasibility zone, the system is kept stable that way, with 0% of outage and with maximum delays that do not exceed the

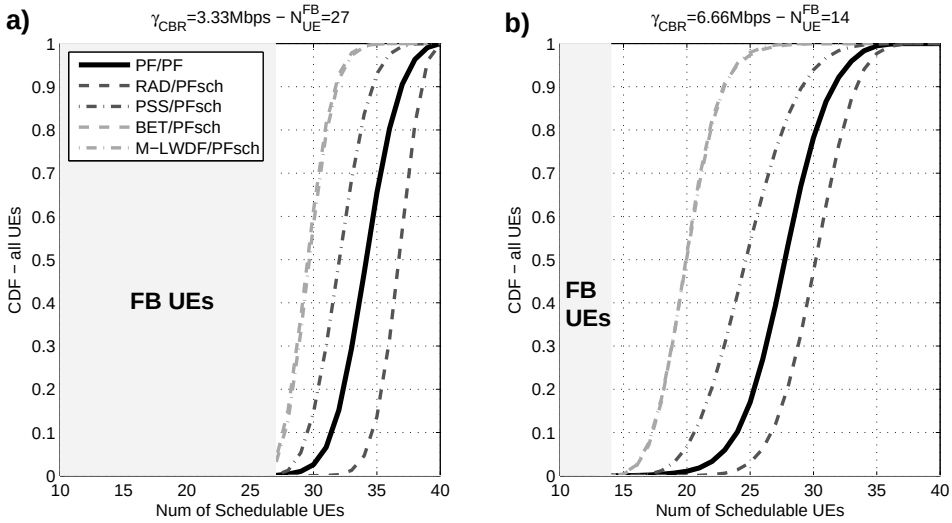

Figure 5.9: CDF of the number of schedulable UEs. The graph (a) shows results for $\gamma = 3.33Mbps$ and $N_{UE}^{BF} = 27$, (b) shows results for $\gamma_{CBR} = 6.66Mbps$ and $N_{UE}^{BF} = 14$. In both graphs, the shaded area corresponds to the number of finite buffer UEs simultaneously in the sector (N_{UE}^{BF}); as finite buffer are always schedulable, the CDF curve of the number of schedulable UEs is always beyond that area.

inter packet arrival time as assessed by Figure 5.8. Note that in those simulation configurations, there is absolutely no difference between BET and M-LWDF regarding the handling of CBR UEs. This behavior is however dependant on the number of CBR UEs. In a situation where $\gamma_{CBR} = 6.66Mbps$ and $CBR = 128kbps$ and therefore 46 CBR UEs present in the cell simulatneously (instead of 26 for $CBR = 256kbps$), the number of schedulable UEs would most likely be double compared to $CBR = 256kbps$ and therefore more often above $N_{mux} = 10$. The two metrics would most likely provide different results.

PSS provides greater delays according to 5.8. The main difference is that no absolute priority is granted to CBR UEs over best effort UEs. Indeed, CBR UEs fall into PSS Set 2 when they comply with their GBR requirement. However, CBR UEs are scheduled according to an approximate maximum throughput rule among each other in Set 2. Indeed, the denominator of the PF metric is nearly equal to the GBR for all CBR UEs. This is a very counter productive behavior which explains the quite higher outage (10%) of PSS and the higher delays.

RAD provides greater delays than BET and M-LWDF as well. The same general reasoning than for PSS applies here too: no explicit absolute priority is granted to CBR UEs over best effort UEs. In a perfect world where the required activity is perfectly estimated, RAD should give results similar to BET and M-LWDF. However, in the imperfect world, $\overline{\mathbf{R_{sch}}}$ is subject to over and underestimations. The overestimations add excessive delays and increase the outage. However, the estimation bias of $\overline{\mathbf{R_{sch}}}$ averages out on a long term and therefore, the final throughput delivered to CBR UEs is close to the GBR.

And the best strategy is... Of course it is the strategy followed by BET and M-LWDF which in the simulations conditions, simply consist in prioritizing systematically CBR UEs over best effort

UEs. Indeed, as mentioned previously, with the power delay profile used in the simulations Typical Urban (TU), the coherence bandwidth is very small compared to the transmission bandwidth. The result is that no gain can be expected from time domain multi-user diversity. Therefore, in those conditions, a CBR UE is served its GBR if it is scheduled a certain number of times, within a certain period. There is therefore only benefit in giving absolute priority to CBR UEs over best effort UEs.

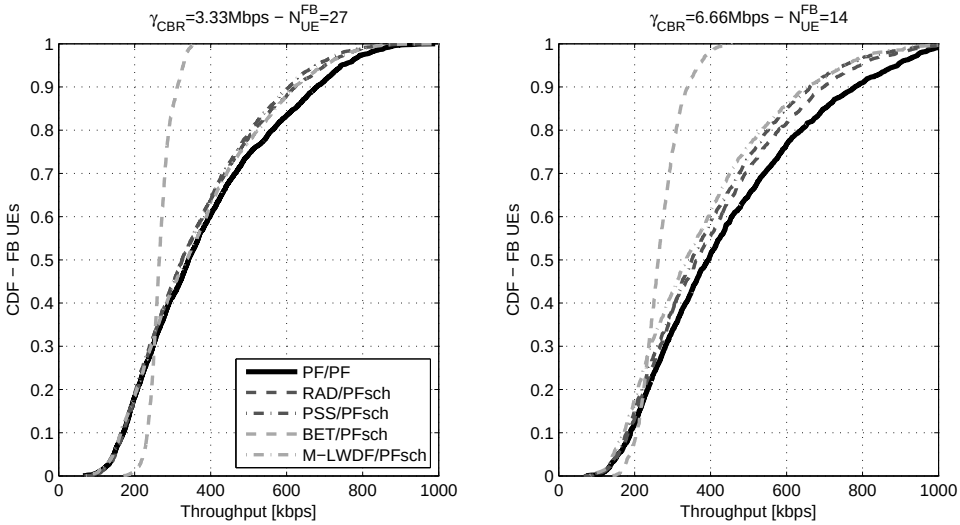

Figure 5.10: CDF of the throughput for finite buffer traffic. The graph (a) shows results for $\gamma_{CBR} = 3.33Mbps$ and $N_{UE}^{BF} = 27$, (b) shows results for $\gamma_{CBR} = 6.66Mbps$ and $N_{UE}^{BF} = 14$

The best effort UEs. Finally, Figure 5.10 shows the CDF of the throughput delivered to finite buffer UEs. It is noticeable that only the BET scheduler proposes a different distribution than the other schedulers. BET follows its natural property and equalizes the different finite buffer UE throughputs. All the other QoS aware algorithms follow a similar PF throughput distribution with less fairness but resulting in a higher average cell throughput.

5.5.1.6 Non Feasible Zone

In the present Section, the behavior of the different algorithms outside the feasible zone, or near its border is analyzed. The two cases with $\gamma_{CBR} = 9.98Mbps$ and $\gamma_{CBR} = 11.77Mbps$ are taken as examples for algorithm analysis. Figure 5.11 shows the UE throughput CDF and Figure 5.12 shows the G-factor versus mean UE throughput for $\gamma_{CBR} = 9.98Mbps$ which is a limit case and $\gamma_{CBR} = 11.77Mbps$ which is clearly out of feasible zone of all algorithms.

PSS and TDPS-BET have equivalent behaviors. This is normal as when γ_{CBR} increases, PSS is not able to provide the GBR to less and less UEs until a point where UEs systematically fall into Set 1 which is strictly equivalent to BET. The behavior of BET can be further analyzed with a simple analysis of the metric. BET schedules the N_{mux} UEs that have an estimated past average throughput $\overline{\mathbf{R}}[t][n]$ below a certain threshold $\mathbf{R}^{th}[t]$. Let us analyze the algorithm with the following simplified assumptions:

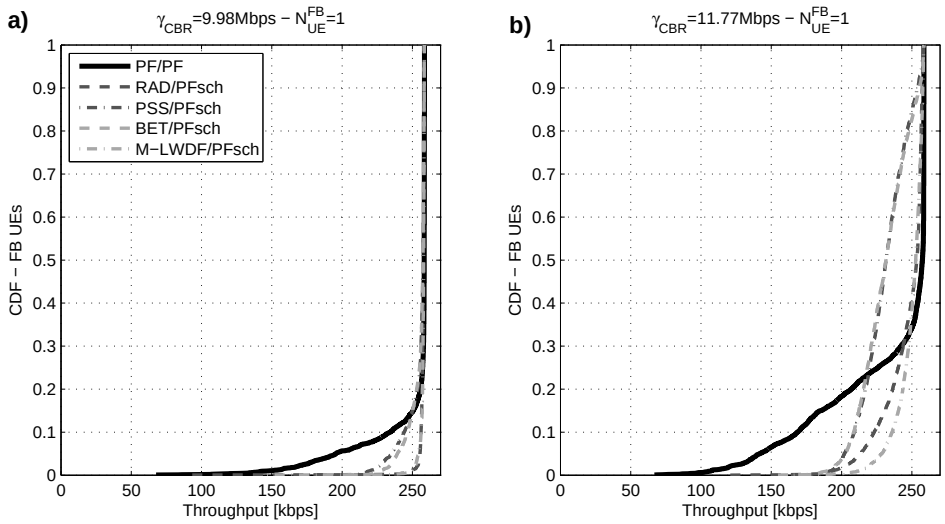

Figure 5.11: UE Throughput CDF for CBR traffic. The graph (a) shows results for $\gamma_{CBR} = 9.98Mbps$, (b) shows results for $\gamma_{CBR} = 11.77Mbps$.

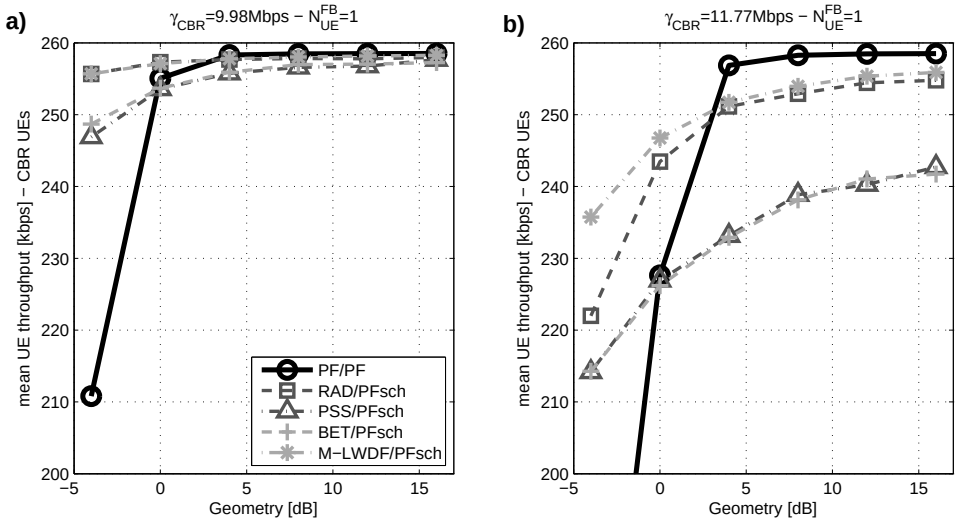

Figure 5.12: G-factor versus average UE Throughput for CBR traffic. The graph (a) shows results for $\gamma_{CBR} = 9.98Mbps$, (b) shows results for $\gamma_{CBR} = 11.77Mbps$.

- *Assumption 1:* When UE n is scheduled, the throughput served is always equal to $\mathbf{Rsch}^a[n]$ for any TTI. In reality, this is of course not true and the served instant throughput will be a more complex process. The served throughput is however expected to average out over time.
- *Assumption 2:* $\mathbf{R}^{th}[t]$ is constant in time. This is not true since:
  - The UEs spatial configuration is not constant over time, therefore, the share of resource that can be given to all UEs will vary when a new UEs comes in the network or when a UE comes out (long term variations)
  - Each estimated UE past average throughput is subject to instantaneous variations (short term variations expected to average out between changes of UEs)

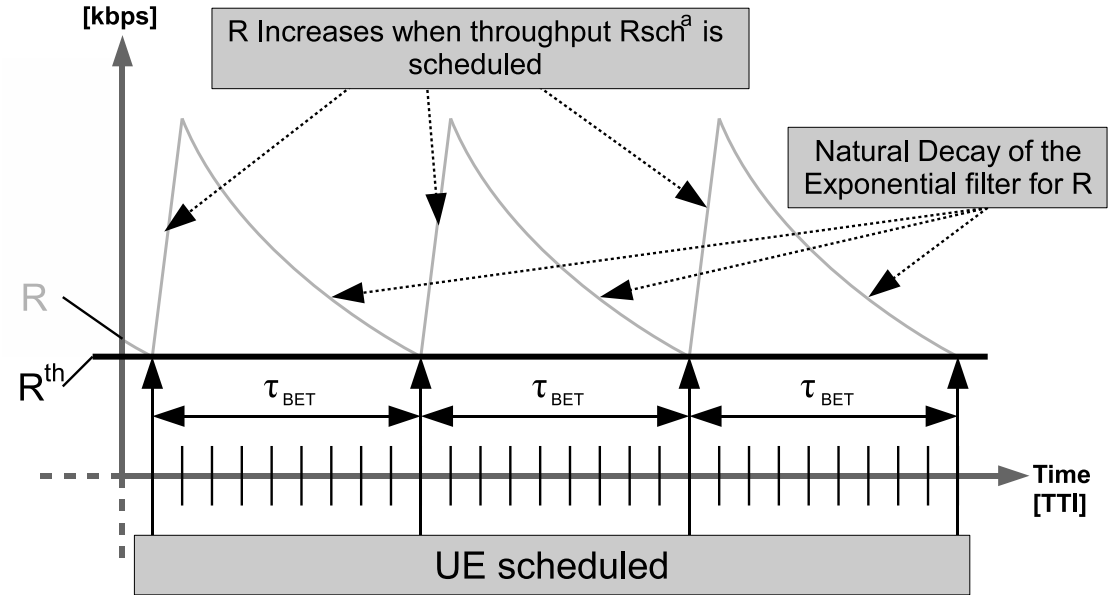

Figure 5.13: Graphical explanation of the simplified assumptions for the analysis of the scheduling process of BET.

Following those assumptions, UE n would be scheduled every $\tau_{BET}[n]$ TTI following the pattern shown in figure 5.13. combining the decay of the exponential filter of the past average throughput $\overline{\mathbf{R}}[t][n]$, considering that UE n is scheduled when $\overline{\mathbf{R}}[t][n] = \mathbf{R}^{th}[t]$ and that the same instant throughput $\mathbf{Rsch}^a[n]$ is always allocated to UE n:

$$\tau_{BET}[n] = \frac{\log\left(1 - \frac{1}{T} \cdot \frac{\mathbf{Rsch}^a[n]}{\mathbf{R}^{th}}\right)}{\log\left(1 - \frac{1}{T}\right)} \tag{5.46}$$

By expanding the numerator and the denominator into Taylor series, $\tau_{BET}[n]$ can be approximated by $\tau_{Id}[n]$ (for 'ideal') where $\tau_{Id}[n]$ is the exact scheduling TTI spacing that should be respected in order to provide throughput $\mathbf{R}^{th}$ to UE n following the assumptions stated in the present Section:

$$\tau_{Id}[n] = \frac{\mathbf{Rsch}^a[n]}{\mathbf{R}^{th}} \tag{5.47}$$

Note that the validity of this approximation increases with T, the constant of the exponential averaging filtering of the UE throughput estimation. Figure 5.14 shows the difference between

τ_{BET} (5.46) and τ_{Id} (5.47) for different values of T and $\mathbf{Rsch}^a[n]$. It is noticeable that τ_{BET} deviates significantly from τ_{Id} for $T = 40$. However, τ_{BET} and τ_{Id} are very similar for $T = 400$ which is the value used in all the simulations.

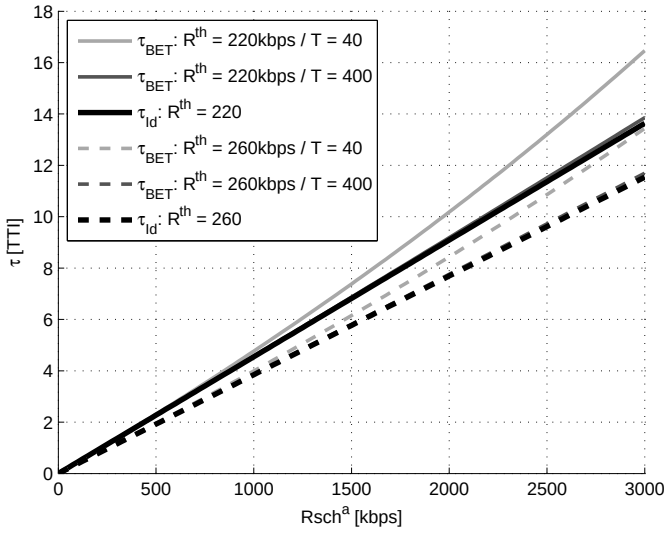

Figure 5.14: $\mathbf{Rsch}^a$ versus τ_{Id} and τ_{BET} for different values of T.

It is concluded that with those simplifying assumptions that,

- BET should provide an equal throughput to all UEs equal to $\mathbf{R}^{th}$;
- Furthermore, the different UEs are served in a pseudo round-robin fashion: UE n is scheduled every $\tau_{BET}[n]$ TTI

In reality, the throughput is not perfectly equalized because *Assumptions* 1 and 2 are not accurate. Furthermore, as the TDPS is selecting N_{mux} UEs every TTI, this allows high Geometry UEs to be provided a higher throughput on average. The result of the BET strategy is that the outage should increases very fast beyond a certain value of γ_{CBR}. Indeed, as the throughput is equalized between the UEs, a situation where $\mathbf{R}^{th} < \alpha \cdot GBR$ would mean that all UEs would be in outage. However, the outage is a hard measure. From a more subjective point of view it is possible to argue that a lot of UEs can stay close to their GBR target. Finally whether BET is the right strategy is matter of policy.

RAD algorithm aims at providing each CBR UEs the exact required time share needed to comply with their GBR requirements. The weighted Round Robin (RR) algorithm is designed so that

- the time share calculated by the RAD component is provided to the UE
- as BET, the different UEs are served in a pseudo round-robin fashion.

However, RAD is expected to scale down the resource share of all UEs compared to the required resource share when the system is not in the feasible zone and behave exactly similarly to BET. However, Figures 5.11 and 5.12 show that this is not the case. The explanation for these

difference lies in the fact that the Weighted Round Robin (WRR) algorithm is a memory-less algorithm. Indeed, when a UE is not scheduled "in time", the counter of WRR is simply reset to 0 forgetting that the UE has undergone a penalty. On the contrary, for BET, when a UE is not scheduled "in time", the past average throughput keep on decreasing.

The result is that RAD is very throughput unfair when it is out of the feasibility zone and gives higher priority to high geometry UEs. Paradoxically, this results in a better scheduling strategy when it comes to outage. Indeed, in the non-feasible zone, in order to maximize the number of satisfied UEs the best strategy is to give a priority according to increasing Geometry.

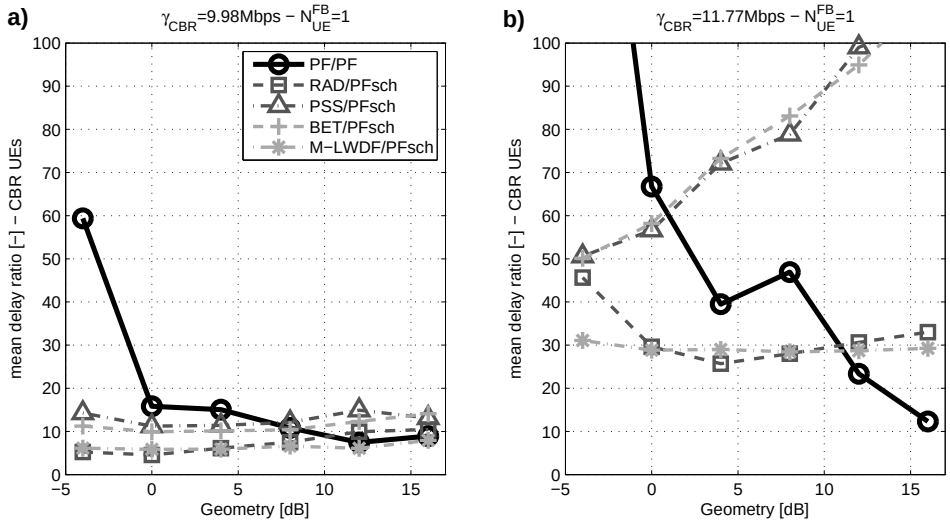

Figure 5.15: G-factor versus average average UE delay ratio for CBR traffic. The graph (a) shows results for $\gamma_{CBR} = 9.98Mbps$, (b) shows results for $\gamma_{CBR} = 11.77Mbps$.

M-LWDF schedules the highest head of line delays weighted with the proportional fair metric. In the simulation conditions:

- the denominator of the proportional fair metric is very close to GBR,
- the numerator, the wideband expected throughput is steady in time, as mentioned before.

Furthermore, as the wideband expected throughput is steady in time, it should also be proportional to the average delivered throughput per TTI defined as:

$$\mathbf{Rsch}^{mean}[n] = \frac{CBR \cdot T_{CBR}}{T_{sch}[n]} \tag{5.48}$$

where T_{sch} is the cumulated time where UE n is scheduled by the TDPS. $\mathbf{Rsch}^{mean}$ is an *a posteriori* measure of the average capacity of the UE. Given this measure, it is possible to determine the minimum average achievable packet delivery delay:

$$\overline{\mathbf{W}_{min}}[n] = \frac{P_{CBR}}{\mathbf{Rsch}^{mean}[n]} \tag{5.49}$$

which corresponds to the serving time of a CBR packet, as if the packet was scheduled every TTI after being received at the eNode-B with a throughput equal to $\mathbf{Rsch}^{mean}[n]$. Finally, the

M-LWDF metric is proportional to $\frac{\Lambda[t][n]}{\overline{\mathbf{W}_{min}[n]}}$ and therefore, should equalize the ratio between the served delay and the minimum average achievable packet delivery delay. In order to verify that, a new statistic is introduced: the delay ratio $\kappa[n]$.

$$\kappa[n] = \frac{\mathbf{W}_{max}[n]}{\overline{\mathbf{W}_{min}[n]}} \tag{5.50}$$

Where $\mathbf{W}_{max}[n]$ is the maximum delay for UE n. Figure 5.15 shows the G-factor versus the average delay ratio. It shows that only M-LWDF equalizes the delay ratio.

Finally, M-LWDF gives generally higher priority to higher geometry UEs, but this priority is regulated by keeping the ratio between the served delay and the minimum average achievable packet delivery delay constant. When γ_{CBR} increases, the proportion of UEs in outage increases and the geometry threshold below which UEs are not satisfied increases.

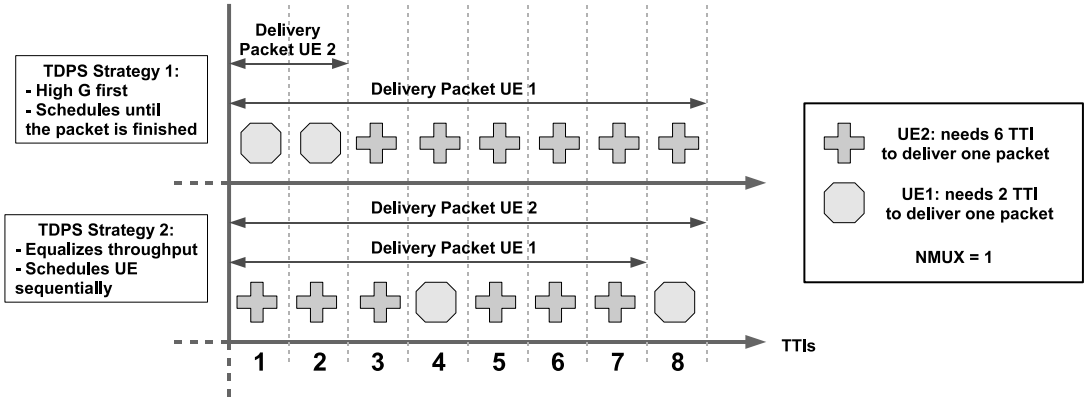

Figure 5.16: Illustration of the effects of two TDPS strategies on packet delivery delay.

Another property of M-LWDF is that while BET and RAD by nature schedule UEs sequentially, M-LWDF will tend to schedule UEs consecutively. Indeed, M-LWDF uses in its metric the head of line delay $\Lambda[t][n]$ which keeps increasing until the packet is delivered. Therefore, the priority metric of a UE keeps increasing until packet delivery. This strategy is beneficial to the overall delay statistics as for equal capacity it is better to completely deliver a packet before starting another transmission. Figure 5.16 illustrates this with a simple example with $N_{MUx} = 1$ and 2 UEs. UE 1 needs only 2 TTIs to deliver one packet while UE 2 needs 6 TTI. In any case, 8 TTIs are needed to deliver one packet of each UE. The best strategy in terms of delay, consists in delivering the packet of UE 1 within the first 2 TTIs and then deliver the packet of UE 2 in the following 6 TTIs. Any other strategy would increase the global delay distribution.

And the best strategy is... It is here a matter of policy. In terms of pure outage, the best algorithms for CBR UEs would be a Max C/I scheduler as it would minimize the number of UEs in outage at the cost of providing no resource to all the UEs that cannot be satisfied. The BET algorithm is the opposite policy that consists in trying to give a part of the resource to everyone knowing that no one will be fully satisfied. Finally M-LWDF is a mid-way alternative that may provide a good balance between proportion of UEs satisfied and overall fairness. The RAD algorithms however needs modifications in order to fulfill its target.

5.5.1.7 Conclusion Regarding The TDPS Scheduling Strategy

Finally, among all the algorithms presented, M-LWDF seems to be collecting the greatest amount of advantages. It performs best in feasible zone and offers an interesting compromise in non feasible zone. Furthermore, it tends to scheduled whole packets at once instead of fragmenting the delivery over time. However, the RAD algorithms in its intention is worth of interest. It is expected that with the following improvements:

- add a compensation component in the WRR metric for delayed or early schedules so that RAD behaves as BET in non feasible zone
- add a delay component to the metric in order to add the possibility to in feasible zone to systematically prioritize CBR UEs over best efforts UEs and to decrease the outage in the non feasible zone;

the RAD algorithms would provide the best packet scheduling framework.

5.5.2 Frequency Domain Scheduling QoS Differentiation

5.5.2.1 Simulation Cases

Table 5.3 summarizes the simulation cases run for the evaluation of different FDPS QoS differentiation PS strategies. The focus is put on the FD-RAD method described in Section 5.4.3. Therefore, the simulation cases involve mixes of traffic where the CBR UEs have a CBR that cannot necessarily be handled by being scheduled every TTI. Different values of γ_{CBR} are tested but the number of CBR UEs is kept constant, equal to 7 while the CBR values are varied from $512kbps$ up to $1536kbps$. Both FDPS-CoItA and FDPS-PFsch are tested with and without FD-RAD as the FDPS scheduler is critical to study here.

5.5.2.2 General Results

The simulations are designed so that most of the outage comes from the inability of the FDPS to provide UE with their GBR when they are scheduled every TTI. Indeed, it has been seen in Section 5.5.1 that values like $\gamma_{CBR} = 3.58$Mbps and $\gamma_{CBR} = 7.17$Mbps can be handled with FDPS-PFsch or FDPS-CoItA and TDPS-RAD while keeping a low outage. However, $\gamma_{CBR} = 10.75$Mbps is near the limit of what throughput a cell can serve and in that case, the outage should come from both the FDPS and the simple fact that the cell is capacity limited. Figure 5.17 shows the outage results of the different cases. It is possible to observe that FD-RAD brings a clear improvement in terms of outage. In the case $\gamma_{CBR} = 10.75$Mbps, FD-RAD even increases the outage. In the following sections, the results are analyzed more in details in order to bring more understanding of FD-RAD.

5.5.2.3 The Effect of FD-RAD on Resource Allocation

The different effects of FD-RAD are summarized in the following Figures. Figure 5.18 shows G-factor versus average throughput of CBR UEs, Figures 5.19 and 5.20 show the G-factor versus the

Table 5.3: Simulation parameters for the evaluation of FDPS RAD algorithms in mixed traffic conditions.

Parameter	**Setting**
Envirronment	MACRO #1
Layout Configuration	1 simulated site
call arrival	N_{UE} constant
	N_{UE} = 20
	T_{FB} = 500s
Traffic Models	Finite Buffer
	N_{UE}^{FB} = 13
	B_{FB} = 2.0Mbits
	GBR_{FB} = 0kbps
	Constant Bit Rate
	N_{UE}^{CBR} = 7
	CBR = 512kbps / 1024kbps / 1536kbps
	T_{CBR} = 10s
	P_{CBR} = 25.6kb
	GBR_{CBR} = 512kbps / 1024kbps / 1536kbps
	γ_{CBR} = 3.58Mbps / 7.17Mbps / 10.75Mbps
TDPS	RAD
FDPS	CoItA with FD-RAD
	CoItA without FD-RAD
	PFsch with FD-RAD
	PFsch without FD-RAD

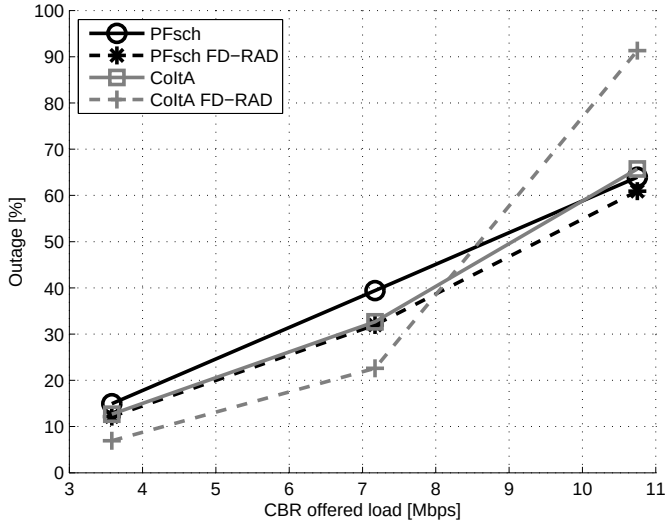

Figure 5.17: Outage values of the different FDPS algorithms associated with FD-RAD or not. The outage is defined by the percentage of CBR UEs with a throughput inferior to their GBR.

average number of PRBs scheduled when the UE is selected by the TDPS for CBR and for finite buffer traffic. Finally, Figure 5.21 shows the throughput CDF of finite buffer UE. Each Figure shows the results of the different algorithms for the different γ_{CBR} cases.

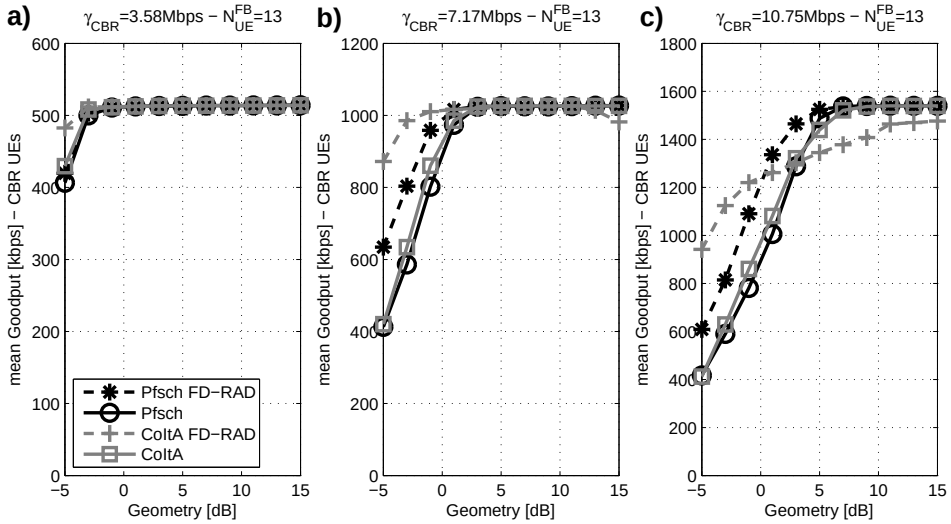

Figure 5.18: G-factor versus average throughput of CBR UEs. The graph (a) shows results for $\gamma_{CBR} = 3.58Mbps$ and $N_{UE}^{BF} = 13$, (b) shows results for $\gamma_{CBR} = 7.17Mbps$ and $N_{UE}^{BF} = 13$ and (c) shows results for $\gamma_{CBR} = 10.75Mbps$ and $N_{UE}^{BF} = 13$.

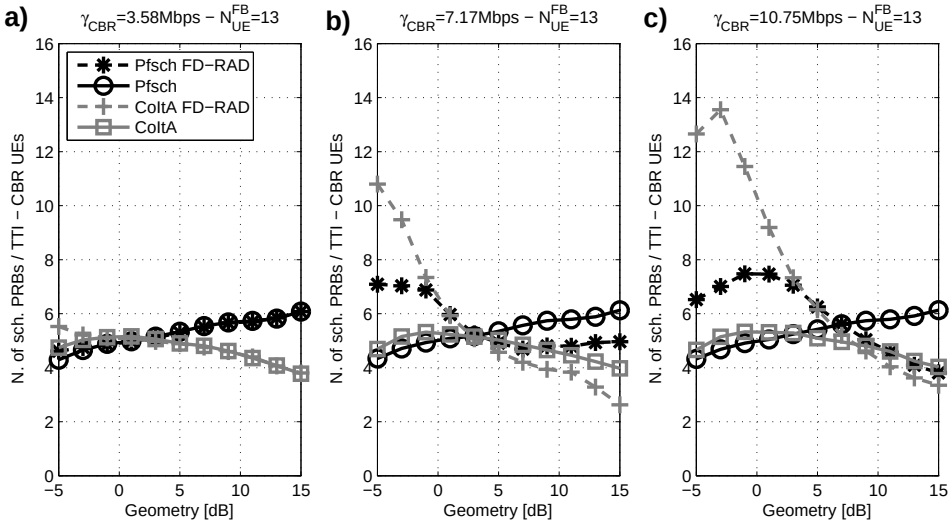

Figure 5.19: G-factor versus average number of scheduled PRBs for CBR UEs. The graph (a) shows results for $\gamma_{CBR} = 3.58Mbps$ and $N_{UE}^{BF} = 13$, (b) shows results for $\gamma_{CBR} = 7.17Mbps$ and $N_{UE}^{BF} = 13$ and (c) shows results for $\gamma_{CBR} = 10.75Mbps$ and $N_{UE}^{BF} = 13$.

Firstly, it is possible to notice that FD-RAD succeeds in separating UEs that are not scheduled enough PRBs from those that are scheduled enough PRBs. Indeed, let us consider as en example the case where $\gamma_{CBR} = 7.17kbps$. For this case, there is a clear threshold for $G = 3dB$. Indeed, below 3db, according to 5.18, the FDPS (PFsch or CoItA) is not able to serve the GBR. But, according to Figure 5.19, when FD-RAD is used, the number of PRBs scheduled to CBR UEs with a G-factor below 3dB is increased. Furthermore, FD-RAD only increases the number of

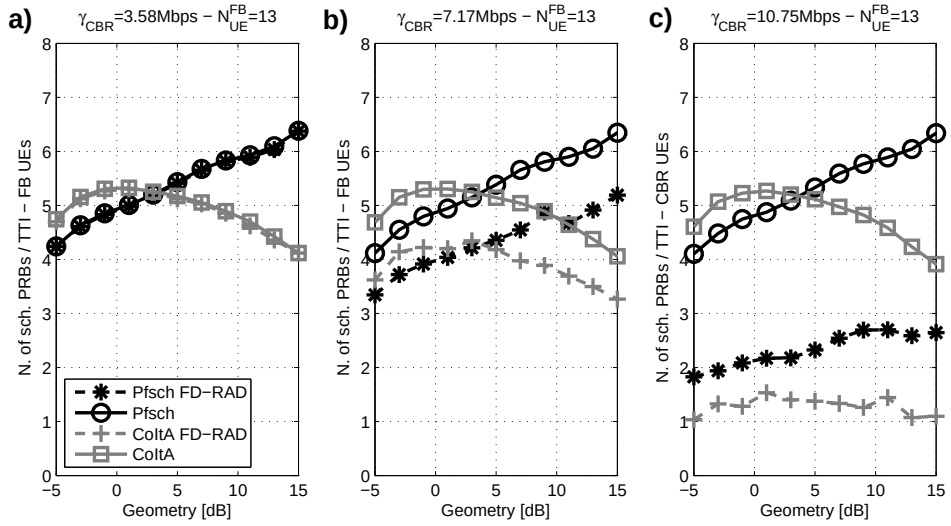

Figure 5.20: G-factor versus average number of scheduled PRBs for finite buffer UEs. The graph (a) shows results for $\gamma_{CBR} = 3.58Mbps$ and $N_{UE}^{BF} = 13$, (b) shows results for $\gamma_{CBR} = 7.17Mbps$ and $N_{UE}^{BF} = 13$ and (c) shows results for $\gamma_{CBR} = 10.75Mbps$ and $N_{UE}^{BF} = 13$.

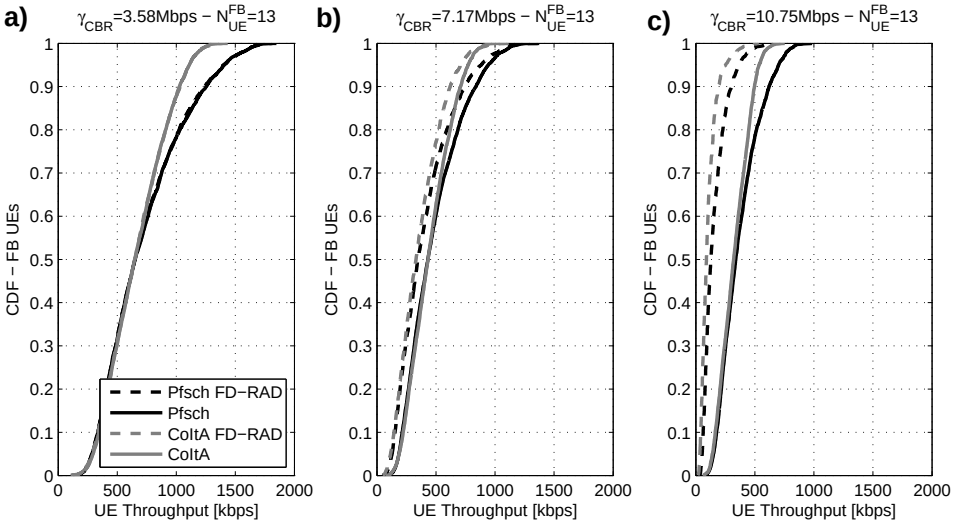

Figure 5.21: Throughput CDF of finite buffer UEs. The graph (a) shows results for $\gamma_{CBR} = 3.58Mbps$ and $N_{UE}^{BF} = 13$, (b) shows results for $\gamma_{CBR} = 7.17Mbps$ and $N_{UE}^{BF} = 13$ and (c) shows results for $\gamma_{CBR} = 10.75Mbps$ and $N_{UE}^{BF} = 13$.

PRBs used for CBR UEs that don't comply with their GBR requirements. Indeed, it is possible to notice on Figure 5.20 that in any case, finite buffer UEs see their number of allocated PRBs decrease when FD-RAD is used. The result is that the throughput served to finite buffer UEs is significantly reduced when FD-RAD is used as show figure 5.21. Similarly, Figure 5.19 shows that number of PRBs allocated to CBR UEs above the G-factor threshold also decreases when FD-RAD is used. Those UEs are simply scheduled more often by the TDPS so that they can comply with their GBR. All those aspects are perfectly consistent with the design of FD-RAD

However, even if FD-RAD succeeds in detecting properly the G-factor threshold, it is not able to provide enough PRBs to lower G-factor UEs in order for them to reach their GBR. This is due to a dimensioning problem of the weighting metric principle. The FDPS metric system in itself guarantees that the number of allocated PRBs increases but it does not guarantee that the given number of PRBs will be enough to serve the GBR. And furthermore, the effect of the metric weighting will be tightly dependent on the FDPS metric itself. Indeed, it is possible to see on Figure 5.19 that for example, FD-RAD combined with CoItA increases more the number of scheduled PRBs than combined with PFsch. The reason is that the effect of metric weighting on PFsch is reduced by a simple contradictory behavior. Adding a weight to the PFsch metric of some UEs tends to increase the metric and the number of scheduled PRBs. However, increasing the number of scheduled PRBs also increases the value of $\mathbf{Rsch}$, which in return decreases the PFsch metric and should decrease the number of scheduled PRBs. Results on Figure 5.19 show that in the end the number of PRBs is still increased with PFsch weighting but significantly less than with CoItA weighting. A simple way to improve the metric weighting principle would be to add a control mechanism insuring that the weight given to a certain UE would increase as long as the UE is not served with a throughput that is enough to guarantee the GBR.

5.5.2.4 Conclusion Regarding The FDPS Scheduling Strategy

Finally FD-RAD works as a throughput increase mechanism. However, it is under dimensionned to actually provide their GBR to all UEs that could not be provided their GBR without FD-RAD. FD-RAD could be improved by improving the metric weighting principle and combining it with a control mechanism insuring that the weighting would be sufficient to guarantee a certain throughput to each UE.

5.6 Conclusion

The present chapter proposes the design of a complete packet scheduling solution for mix of best effort and real time traffic. Based on the conclusions of Chapter 3, the Required Activity Detection concept is developed into both time and frequency domain schedulers. Furthermore, some algorithms from the literature are adapted to support mix of traffics.

A detailed analysis of the time domain schedulers through extensive simulations led to the following conclusions. Firstly The Modified Largest Weighted Delay First (M-LWDF) has the best overall performance and is the preferred time domain scheduler in the simulated conditions. Indeed, in the feasibility zone, M-LWDF gives absolute priority to real time UEs resulting in an outage of 0% in that zone. In the non-feasibility zone, M-LWDF maintains the system in an equilibrium which is a compromise between the throughput fairness strategy (followed by Blind Equal Throughput and that yelds a very high outage) and the maximum C/I strategy (followed by

the PF scheduler in the non feasibility zone and that yelds the lowest outage, hiding the fact that many users are simply never scheduled).

The Time-Domain Required Activity Detection (TD-RAD) is not recommended for implementation as derived in the thesis. However, The TD-RAD would present several advantages in a more complex scenario. Indeed, for example, the RAD concept makes it easy to control the time share given to best effort users. Therefore, based the learnings of the study, it would be beneficial to improve the TD-RAD algorithm by adding a delay component to the metric in order to combine the different qualities of M-LWDF and of the RAD concept.

Finally, the Frequency-Domain Required Activity Detection (FD-RAD) simply applies weights to the Frequency-Domain Packet Scheduling metric and shows to greatly improve the system performance when high GBR UEs are present in the system. In the results, FD-RAD can decrease the outage of up to 50%. Therefore, FD-RAD is recommanced for implementation in a real system on top of the existing Frequency-Domain Packet Scheduling metric.

Chapter 6

Radio Admission Control

6.1 Introduction

Radio Admission Control (RAC) is the mechanism that aims at keeping the cell in a state where the outage is low by admitting or blocking User Equipment (UE)s incoming in network. The RAC algorithms are evaluated in terms of UE unsatisfaction where the unsatisfied UEs include:

- The UEs in outage as in the Packet Scheduler (PS) studies,
- And the UEs blocked by the RAC.

In this Chapter, state of the art RAC schemes are compared with a new Required Activity Detection (RAD) based RAC algorithm. The study takes into account the new possibilities included in Long Term Evolution (LTE) compared to earlier 3rd Generation Partnership Project (3GPP) releases. The most important change lays in the possibility for the RAC to access Layer 2 information as the RAC is located in the E-UTRAN Node B (eNode-B) while for example, in High-Speed Downlink Packet Access (HSDPA), the RAC was located in another node called the Radio Network Controler (RNC). The rest of the Chapter is organized as follows: Section 6.2 describes RAC framework used in the thesis work. Then Sections 6.3 and 6.4 describes in details throughput and RAD based RAC. Finally, the performance of those algorithms is studied in Section 6.5 and the Chapter is concluded in Section 6.6.

6.2 RAC framework

In the present study, the RAC will be formalized in the following way. The state of the system is characterized by a one dimensional unit value $\mathbf{K_{cell}}\left[t\right]$ where each UE brings a linear contribution $\mathbf{K_{UE}}\left[t\right]\left[n\right]$ so that:

$$\mathbf{K_{cell}}\left[t\right] = \sum_{n=1}^{N_{UE}} \mathbf{K_{UE}}\left[t\right]\left[n\right] \tag{6.1}$$

The unit of $\mathbf{K_{cell}}\left[t\right]$ can be the throughput (expressed in Mbps for example), the required transmission power (expressed in dBm for example), the number of UEs or any other unit that helps

quantifying the state of the system in terms of resource need. Besides, the feasibility zone is characterized by a threshold value K_th that corresponds to the upper bound of the resource that the system can provide. Therefore, the criterion for accepting a new UE in the network is:

$$\mathbf{K_{cell}}[t] + \mathbf{K_{new}} < K_{th} \cdot (1 - \Delta) \tag{6.2}$$

where $\mathbf{K_{new}}$ is the resource need of the incoming UE. Δ represents a safety margin. If the algorithm is too permissive by nature, using a positive Δ will help keeping the system in the feasibility zone. On the contrary, if the algorithm is to conservative by nature, using a negative Δ will increase the number of accepted UEs and therefore decrease the overall unsatisfaction.

Besides, figure 2.3 of Section 2.4 shows the LTE modeling of RAC used in the thesis. The RAC functionality can use all the elements that the PS is using. This fact is a novelty in recent 3GPP releases. The RAC can therefore use the Channel Quality Information (CQI) of the UEs that are present in the network as well as the same variable used for the PS algorithms. However, an incoming UE usually does not have any available CQI. The RAC can instead use Layer 3 channel quality indications like the Reference Signal Received Quality (RSRQ), which consists of the ratio between the received reference signal and the received signal on a certain section of the bandwidth. The RSRQ is fully defined in [60].

6.3 Throughput Based Radio Admission Control

The throughput based algorithms characterize the state of the system with the capacity. Two throughput based algorithms are described in the present section. The first algorithm called *fixed throughput based RAC* is a very simple algorithm that is used as a reference when evaluating the performance of the different algorithms. The second algorithms called *adaptive throughput based RAC* is a more advanced scheme based on [51].

6.3.1 Fixed Throughput Based RAC

Throughput based algorithms follow the following logical questioning:

1. What is the maximum throughput of the cell?

In the fixed throughput algorithm, the maximum cell throughput is assumed to be fixed and equal to C_{th}, a value chosen as parameter. This assumption does not correspond to reality since the cell throughput changes with the channel quality of UEs present in the cell, with the Quality of Service (QoS) parameters and also depends on the PS algorithms employed. It also changes depending on the interference level from neighboring cells. However, assuming that the capacity is constant is still reasonable since a cell has characteristic physical conditions (for example: the physical coverage area, the transmission bandwidth) that do not change in a long term. Therefore, assuming the interference level from neighboring cells constant, each cell should have an average capacity even though the UE locations are changing in time. As an example, the baseline simulations in Chapter 2 show that the system in the simulated conditions has a throughput of around 10.5Mbps.

2. What is the capacity occupied by the UEs present in the cell?

It is assumed here that the packet scheduling algorithm is providing to each UE its Guaranteed Bit Rate (GBR). Therefore, the capacity required by the UEs present in the cell is equal to the sum of the different UEs GBR.

3. What capacity is the incoming UE requiring?

Following the same assumption, the capacity required by the incoming UE is equal to its GBR. Finally, the fixed throughput algorithm is summarized by (6.3):

$$\sum_{n=1}^{N_{UE}} GBR_n + GBR_{new} < C_{th} \tag{6.3}$$

If the expression is true then the incoming UE is accepted, otherwise, the UE is rejected. This algorithm is very simple to implement and to apply to a system. However, it has two major disadvantages. Firstly, the capacity threshold C_{th} must be manually set. The operator must therefore have a precise idea of the possible capacity in the system. Furthermore, the capacity even if average should oscillate depending on the UE configuration present in the system. For example, the capacity will decrease when a high geometry UE is replaced by a low geometry UE.

6.3.2 Adaptive Throughput Based RAC

In order to solve the disadvantages of the fixed throughput algorithms, the adaptive throughput has been developed. The adaptive throughput based algorithm follows exactly the same principle than the fixed throughput algorithms. The difference lies in the estimate of the cell capacity. The adaptive throughput algorithm tries to estimate dynamically the capacity of the cell based on the past cell throughput. The adaptive throughput algorithm is summarized in (6.4) where $\overline{\mathbf{C_{th}}}[t]$ is the dynamically estimated capacity at Transmission Time Interval (TTI) t.

$$\sum_{n=1}^{N_{UE}} GBR_n + GBR_{new} < \overline{\mathbf{C_{th}}}[t] \cdot (1 - \Delta) \tag{6.4}$$

The maximum supported cell throughput is estimated with a simple exponential filtering of the scheduled throughput as described in (6.5):

$$\overline{\mathbf{C_{th}}}[t] = \min\left(\max\left(\frac{T_C - 1}{T_C} \cdot \overline{\mathbf{C_{th}}}[t-1] + \frac{1}{T_C} \cdot \mathbf{C_{th}}[t], C_{min}\right), C_{max}\right) \tag{6.5}$$

Where $\mathbf{C_{th}}[t]$ is the instant capacity estimate of the cell. Note that an upper bound C_{max} and a lower bound C_{min} are introduce to prevent the algorithm of estimating the capacity at a too high or too low level. When the number of UEs is very low, the capacity may be estimated from a non representative number of UEs, therefore, in this case, the capacity may be either underestimated (for example if there is only one UE in the cell, at the cell edge) or overestimated (for example if there is only one UE in the cell, close to the eNode-B). The instant capacity is estimated as follows:

$$\mathbf{C_{th}}[t] = \frac{N_{PRB}}{\mathbf{N_{PRB}^{sch}}[t]} \cdot \mathbf{C_{inst}}[t](1 - BLER) \tag{6.6}$$

Where $\mathbf{C_{inst}}[t]$ is the scheduled cell throughput at TTI t, $\mathbf{N_{PRB}^{sch}}[t]$ the number of Physical Resource Block (PRB)s scheduled at TTI t and N_{PRB} the total number of PRBs. $\mathbf{C_{inst}}[t]$ is scaled

with the $BLER$ and with the ratio between the total number of PRBs and the number of PRBs that have been used.

The adaptive throughput algorithm aims at adapting the cell throughput estimate to the variation of UEs configurations in the network. The disadvantage of this algorithm is that it does not take into account the incoming UEs channel quality and therefore relies on a minimal change of the cell capacity once the UE has entered the network. The precision of the algorithms will therefore increase with the number of UEs present in the network. However, one advantage of this algorithm is that the capacity estimate is based on the past schedules and therefore takes into account the PS influence on the cell capacity as well as neighboring cells interference.

6.4 Required Activity Detection Based Radio Admission Control

The RAD based RAC is a new RAC scheme where the state of the system is characterized by the required activity. The required activity here is defined by the average number of PRBs per TTI needed to serve a UE with its GBR. In the principle, the RAD based algorithm corresponds to a similar logical questioning than the previous algorithms:

1. What is the total available number of PRBs per TTI?

It is a fixed value equal to $N_{PRB} \cdot (1 - BLER)$ as it is assumed that a fraction of the resource equal to $(1 - BLER)$ is used for retransmissions.

2. What share of the total required activity is required by the already existing UEs?

For each UE n, the PS can calculate the past average scheduled throughput per PRB: $\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}[n]$. The definition of this variable is given in Section 5.4.3. This value gives an estimate of the achievable throughput per PRB. Therefore, the required activity of UE n can be estimated as:

$$\mathbf{RA}[t][n] = \min\left(\frac{GBR_n}{\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}[n]}, RA_{max}\right). \tag{6.7}$$

where RA_{max} is a parameter allowing to set a maximum required activity per UE. It is necessary if the required activity is estimated too high in order to keep on accepting UEs in the network. The reason is that $\mathbf{R}_{\mathbf{sch}}^{\mathbf{prb}}[n]$ is calculated by exponential averaging and can therefore reach very small (and unrealistic) values at times. The total required time-frequency scheduling fraction can be calculated by simply summing over all the UEs:

$$\mathbf{RA}_{\mathbf{cell}}[t] = \sum_{n=1}^{N_{UE}} \mathbf{RA}[t][n] \tag{6.8}$$

3. What frequency-time scheduling fraction is required by the incoming UE?

The required frequency-time scheduling fraction of the incoming UE can be estimated by the following ratio:

$$RA_{new} = \frac{GBR_{new}}{\hat{D}_{new}} \tag{6.9}$$

Where $\hat{D}_{new}$ is an estimate of the throughput of the incoming UE if it was given all the bandwidth. This value can be estimated by using the modified Shannon formula derived in [75]:

$$\hat{D}_{new} = Bdw \cdot \alpha_{eff} \cdot \log_2 \left(1 + \hat{G}_{new} \cdot A_{new}\right) \tag{6.10}$$

Where:

- $\hat{G}_{new}$ is the G-factor estimation of the incoming UEobtained from the RSRQ (details about the RSRQ and the geometry estimation are given in a later stage),
- Bdw is the transmission bandwidth,
- α_{eff} is the proportion of the bandwidth used efficiently. Taking into account the different efficiency loss mainly due to the diverse signaling and reference signaling, it has bee shown in [75] that the bandwidth efficiency of LTE is of 0.53 for $Bdw = 10$MHz.
- A_{new} is the array gain which depends mainly on the number of antennas. The simulations are run with Maximal Ratio Combining (MRC) with 1 transmit antenna and 2 receive antennas. Therefore we simply take $A_{new} = 3$dB

Finally, the criterion for accepting the incoming UE for the RAD based RAC algorithm is:

$$\mathbf{RA_{cell}}\,[t] + RA_{new} < N_{PRB} \cdot (1 - BLER)\,. \tag{6.11}$$

The RAD based RAC has the same advantages than the adaptive capacity based RAC: it adapts to the different conditions of the system depending on the configuration of the UEs. Moreover, RAD based RAC takes into account the channel quality of the incoming UE which is not the case of capacity based RAC algorithms. Therefore, the accuracy of the RAD based RAC does not depend on the number of UEs in the network.

6.5 Performance Evaluation

6.5.1 Geometry evaluation modeling

In the simulations carried out in the thesis, the RSRQ calculation is not explicitly implemented, therefore, geometry estimation model that includes the characteristics of the RSRQ measurements is built. L3 measurements comprise RSRQ, Reference Signal Received Power (RSRP) and Reference Signal Strength Indicator (RSSI) defined as follows [60]:

- RSRP is a linear average over the power contributions (in [W]) of the resource elements that carry cell-specific reference signals within the considered measurement frequency bandwidth. It includes antenna pattern, antenna gain, path gain and slow fading.
- RSSI is the total received power observed by the UE from all sources, including co-channel serving and non-serving cells, adjacent channel interference, thermal noise, etc.
- RSRQ is simply the ratio between RSRP and RSSI. Note that RSRP and RSSI are measured on the same bandwidth:

$$RSRQ = \frac{RSRP}{RSSI} \tag{6.12}$$

According to the definition, the RSRQ can be taken directly as geometry estimate. However, it is important to be aware of certain characteristics:

- *ch1:* RSRP measures only the reference signaling while RSSI contains both reference signaling and data signaling. Therefore, there will be we shift between geometry and RSRQ due to the difference in transmitted reference signal power and transmitted signaling power. RSRQ will be significantly lower than the geometry. In the study it is assumed that this shift is known and that it is constant since the system operates at full load.

- *ch2:* RSSI measures the global signal strength including the signal strength from the own cell. It means that for geometry values above 0, the RSRQ is not reliable as an estimate of the geometry and will have an almost constant value.

- *ch3:* RSRQ is subject to measurement errors due to thermal noise and fast fading averaging.

Therefore, a simple geometry estimation model is built:

$$\hat{G} = \min(0, G) + \epsilon_{1dB} \tag{6.13}$$

where ϵ_{1dB} is a normally distributed random generated variable centered in 0 with a standard deviation of 1dB that simulates the measurement errors due to thermal noise. Furthermore, all the geometry values above 0 are simply on average 0 due to *ch2*. Indeed, while the geometry can be expressed as follow:

$$G = \frac{P_{ownCell}}{P_{otherCells} + P_{noise}} \tag{6.14}$$

while the RSRQ can be expressed as follows:

$$RSRQ = \frac{P_{ownCell}}{P_{ownCell} + P_{otherCells} + P_{noise}} \tag{6.15}$$

and therefore $RSRQ$ converges toward 1 (thus 0dB) as the own cell power increases. Figure 6.1 gives an example of geometry versus geometry estimates.

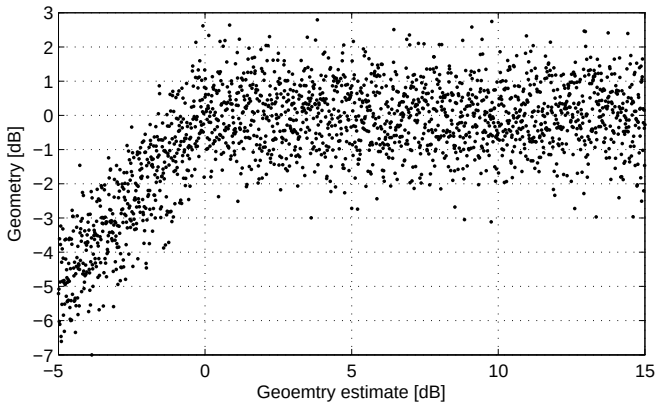

Figure 6.1: Examples of geometry versus geometry estimates from RSRQ

Table 6.1: Simulation parameters for the evaluation of RAC algorithms.

Parameter	**Setting**
Envirronment	MACRO #1
Layout Configuration	1 simulated site
call arrival	Poisson
Traffic Models	Finite Buffer
	$\overline{\gamma_{FB}}$ = 2Mbps
	B_{FB} = 2.0Mbits
	GBR_{FB} = 0kbps
	Constant Bit Rate
	$\overline{\gamma_{CBR}}$ = 7.5Mbps / 9Mbps / 10.5Mbps / 12Mbps
	CBR = 256kbps
	T_{CBR} = 10s
	P_{CBR} = 25.6kb
	GBR = 256kbps
TDPS	RAD
FDPS	PFsch
RAC	Fixed Throughput
	C_{th} = 10.3Mbps
	= 11.8Mbps
	= 13.3Mbps
RAC	Adaptive Throughput
	C_{min} = 8Mbps
	C_{max} = 15Mbps
	Δ = -0.1
	= -0.2
	= -0.3
	= -0.4
RAC	RAD
	RA_{max} = 0.1
	Δ = -0.1
	= -0.2
	= -0.3
	= -0.4
	$\hat{G}$ = $\min(0, G) + \epsilon_{1dB}$
	= 0

6.5.2 Simulation Cases

The different simulation cases run to evaluate the performance of the RAC algorithms are summarized in Table 6.1. The performance evaluation is performed with *a mix of Constant Bit Rate (CBR) UEs and finite buffer UE* with Poisson call arrival. The average offered load of finite buffer UEs is kept constant: $\overline{\gamma_{FB}} = 2Mbps$. Three different CBR offered load values are tested: $\overline{\gamma_{CBR}} = 9Mbps$, $\overline{\gamma_{CBR}} = 10.5Mbps$ and $\overline{\gamma_{CBR}} = 12Mbps$, thus covering cases where the CBR offered load is below the cell capacity, approximately equal to the cell capacity and largely above cell capacity. The fixed throughput algorithm is tested with $C_{th} = 10.3$, 11.8 and 13.3. The RAD algorithm is tested with both

- the geometry estimation of incoming UEs described in Section 6.5.1
- geometry estimated as being 0dB for all incoming UEs.

The goal of this later case is to evaluate the influence of the geometry estimation on the RAD algorithm. Indeed, the RAD algorithm has two mechanisms that can help improving the overall UE satisfaction: the channel based required activity estimation and the possibility of selecting incoming UEs depending on their channel quality.

6.5.3 Overall Performance

The main evaluation criterion of RAC algorithms is the satisfaction rate P_{sat}. The satisfaction rate is defined by the proportion of UEs that are neither blocked nor in outage:

$$P_{sat} = (1 - P_{block}) \cdot (1 - P_{outage}) \tag{6.16}$$

where P_{block} is the proportion of UE that are blocked among all calls and P_{outage} is the proportion of UEs in outage among the UEs that are accepted in the network. The unsatisfaction rate P_{uns} is the complementary of the satisfaction rate:

$$P_{uns} = 1 - P_{sat} \tag{6.17}$$

Figure 6.2 summarizes the overall performance of the different RAC algorithms. The reference RAC is the fixed throughput based with $C_{th} = 18Mbps$. C_{th} is set to a high value so that the RAC has a very low blocking rate. Otherwise, on Figure 6.2 only one special tuning (one value of Δ or one value of C_{th}) of each of the four available algorithms is presented in Figure 6.2. It is chosen to show the tunning that minimizes the unsatisfaction rate for $\overline{\gamma_{CBR}}$ = 7.5Mbps and $\overline{\gamma_{CBR}}$ = 9Mbps. Indeed, in a real system, a cell should be dimensioned so that the blocking rate does not exceed a few percents in order to provide a sufficient satisfaction to the UEs. Hence in the simulation settings, only $\overline{\gamma_{CBR}}$ = 7.5Mbps and $\overline{\gamma_{CBR}}$ = 9Mbps can provide blocking rates that are below 5%. Higher offered load values are less relevant in the sence that they may not occur in a real system. However, higher offered load cases are interesting from a performance analysis point of view.

The overall results show the effect and the importance of the RAC. Indeed, with $C_{th} = 18Mbps$, the unsatisfaction rate is very high for all the simulation cases. The other algorithms provide a reduction of the order of 30% in unsatisfaction rate at $\overline{\gamma_{CBR}} = 9Mbps$ and of the order of 50% at $\overline{\gamma_{CBR}} = 10.5Mbps$. There is only a small difference between the other RAC algorithms (excluding reference) for low loads with an unsatisfaction rate of 2.5% for $\overline{\gamma_{CBR}} = 7.5Mbps$ and

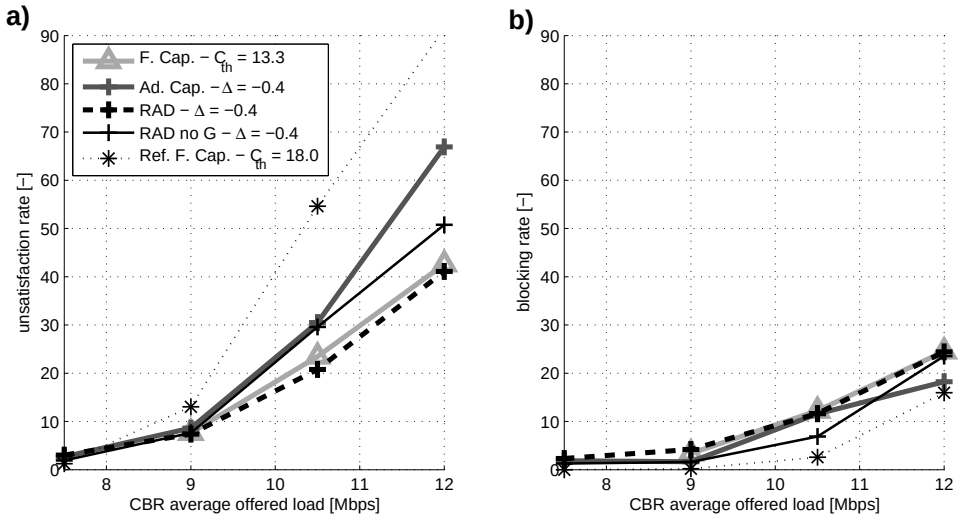

Figure 6.2: Overall performance of the different RAC algorithm with best parameterization. Figure (a) shows the unsatisfaction rate and Figure (b) shows the blocking rate

an unsatisfaction rate of approximately 7.5% for $\overline{\gamma_{CBR}} = 9Mbps$. However, for higher values, the RAD-based RAC outperforms all the other algorithms. Fixed capacity algorithms show a slightly inferior performance than RAD while RAD with no geometry information and adaptive capacity show the worst results.

6.5.4 The Fixed Capacity RAC Algorithms

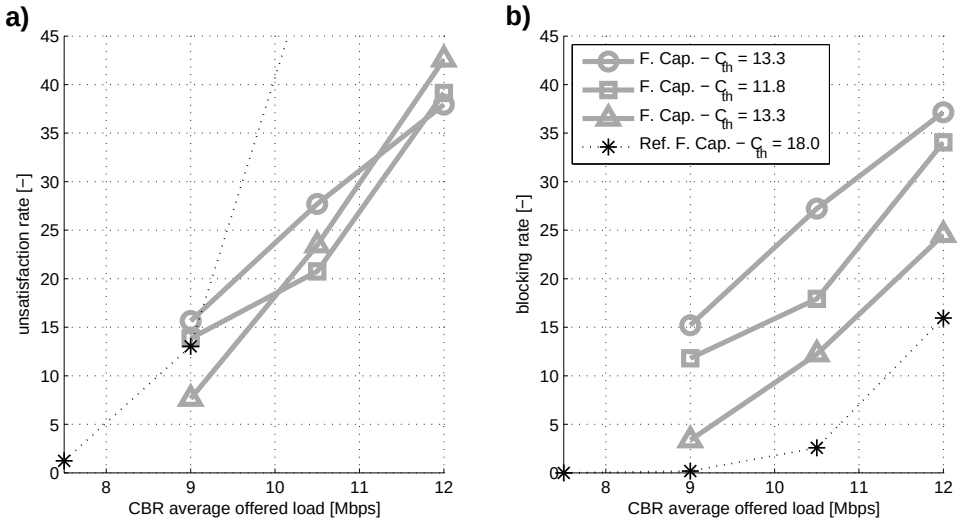

Figure 6.3: Overall performance of the fixed capacity algorithm. Figure (a) shows the unsatisfaction rate and Figure (b) shows the blocking rate

Figure 6.3 shows the results of the fixed capacity RAC algorithms for different values of C_{th}. The different values of C_{th} show different compromise between blocking and outage. For C_{th} = 10.3Mbps and C_{th} = 11.8Mbps, the unsatisfaction rate (a) and the blocking rate (b) are similar,

which means that the outage is very low. Therefore those two values keep the system in the feasibility zone. On the contrary, with C_{th} = 13.3Mbps, the outage increases significantly with $\overline{\gamma_{CBR}}$, to the point that for $\overline{\gamma_{CBR}}$ = 10.5Mbps and $\overline{\gamma_{CBR}}$ = 12Mbps, C_{th} = 11.8Mbps outperforms C_{th} = 13.3Mbps.

Two different RAC strategies are here highlighted:

- a *hard* strategy where the aim is to keep the outage close to 0% and the system remains therefore strictly in the feasibility zone. This is the case with C_{th} = 10.3Mbps and C_{th} = 11.8Mbps.
- a *soft* strategy where the outage can be compromised in order to improve the overall UE satisfaction. This is the case with C_{th} = 13.3Mbps.

For low $\overline{\gamma_{CBR}}$, the *soft* strategy seems to be more appropriate as the outage remains low. However, for higher loads, the outage increases too much, therefore, the *hard* strategy is more appropriate. When using a *soft* strategy, the RAC artificially considers limits beyond the actual feasibility zone. With a low offered load, the chances that the system goes beyond the limits of the feasibility zone are low so that when it happens the time spend by the system outside the feasibility zone is small, thus creating only small penalties in terms of outage. However, when the offered load is high, the system spends a long time beyond the feasibility zone. When the system is outside the feasibility zone, all UEs are affected and the outage increases.

It is important to note with the fixed capacity RAC algorithm, the tunning is totally dependent on the cell environment (3GPP MACRO#1 cell in the present case). In a real network, every cell is potentially different and has different physical characteristics. Therefore, the fixed capacity algorithm my not be suited for a real case application.

6.5.5 The Adaptive Capacity RAC Algorithm

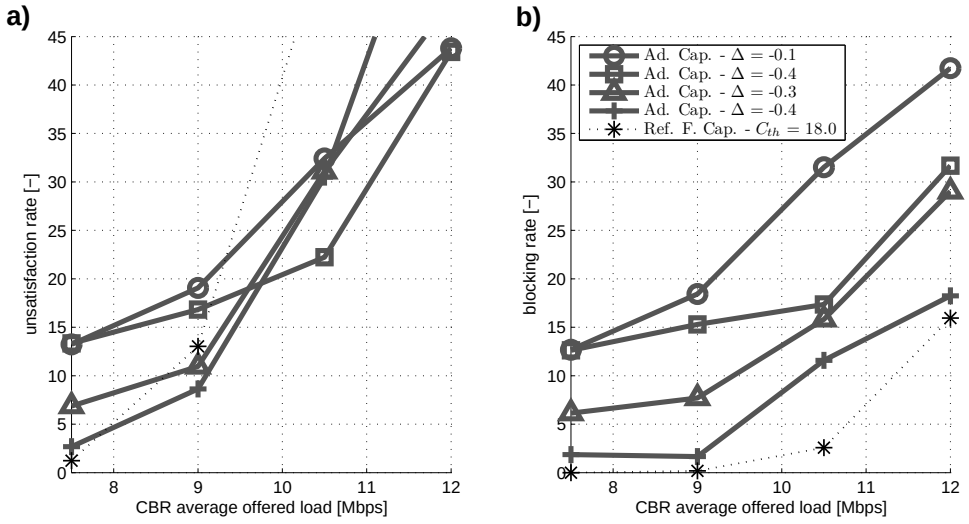

Figure 6.4: Overall performance of the fixed capacity algorithm. Figure (a) shows the unsatisfaction rate and Figure (b) shows the blocking rate

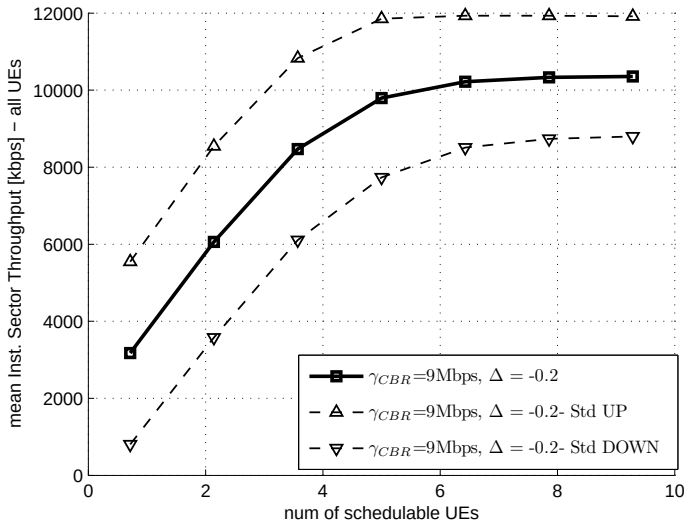

Figure 6.5: Number of scheduled UEs per TTI versus average instant cell throughput

The performance of the adaptive throughput algorithm is investigated in the present Section. The performance results are summarized in Figure 6.4. Firstly, figure 6.5 shows the number of scheduled UEs per TTI versus the average instant cell throughput. It clearly appears that the instant throughput decreases significantly when the number of schedulable UEs is below 5. This is due to the low geometry finite buffer UEs that stay longer in the network since finite buffer are provided equal share of resource by the Time Domain Packet Scheduling (TDPS)-RAD algorithm.

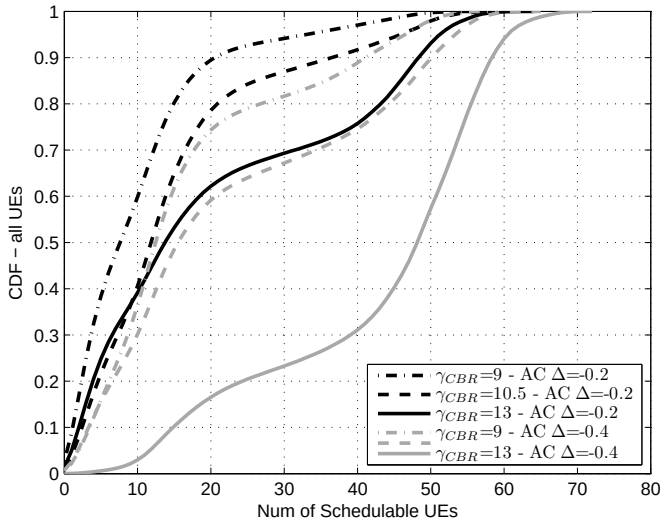

Figure 6.6: Number of schedulable UEs per TTI CDF for the adaptive capacity algorithms

Besides, Figure 6.6 shows the CDF of the number of schedulable UEs per TTI. It appears that the number of schedulable UEs is often below 5. This results therefore in an instant throughput that is often very low as shown in the instant cell throughput CDF on Figure 6.7. The consequence is that the adaptive capacity algorithm is too conservative and as seen in Section 6.5.3, needs to be compensated with a negative Δ value in order to start showing UEs in outage. More importantly,

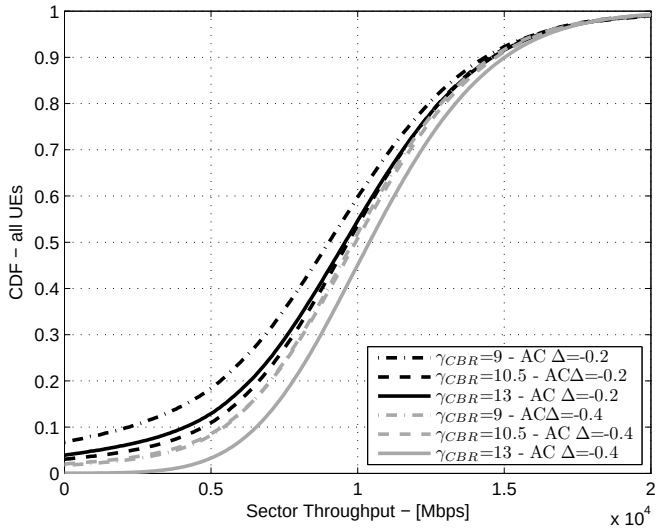

Figure 6.7: Instantaneous cell throughput CDF for the adaptive capacity algorithms

it is shown here that the cell capacity estimate can be built from situations where only a few low geometry finite buffer are present in the cell, which is of course not representative of the real capacity of the cell and simply introducing a shift in this capacity estimate does not make it more precise than a simple fixed value but simply closer to the real value. It can be argued that the type of situation with a low number of schedulable UEs is very specific to the simulation conditions. Indeed, in a situation with more UEs with a lower CBR, the number of schedulable UEs would be higher and the problem would appear less and the adaptive capacity algorithms would probably yield better results in such a situation. However, the simple fact that the algorithms shows poor results in a situation that is realistic is enough to contraindicate the adaptive capacity RAC algorithm.

6.5.6 The RAD algorithm

The RAD-based RAC algorithm is simulated under two different forms as mentioned in Table 6.1. Firstly, Figure 6.8 summarizes the results of the RAD-based RAC algorithm where the geometry information for incoming UEs is taken into account under the modeling expressed in (6.13). Then, Figure 6.9 shows the results of the RAD-based RAC algorithms where the geometry of incoming UEs is systematically estimated to be 0. With those two sets of simulations, it is possible to evaluate the influence of taking into account the geometry information of incoming UEs, under the model depicted in Section 6.5.1. It is important to notice that the same effect than for fixed capacity and adaptive capacity is observed for the RAC-based RAD algorithm. Indeed, a *soft* strategy (higher Δ) seems to have a better payoff for lower offered loads (7.5Mbps and 9Mbps) while a *hard* strategy (lower Δ) seems to have a better payoff at higher offered loads (10.5Mbps and 12Mbps).

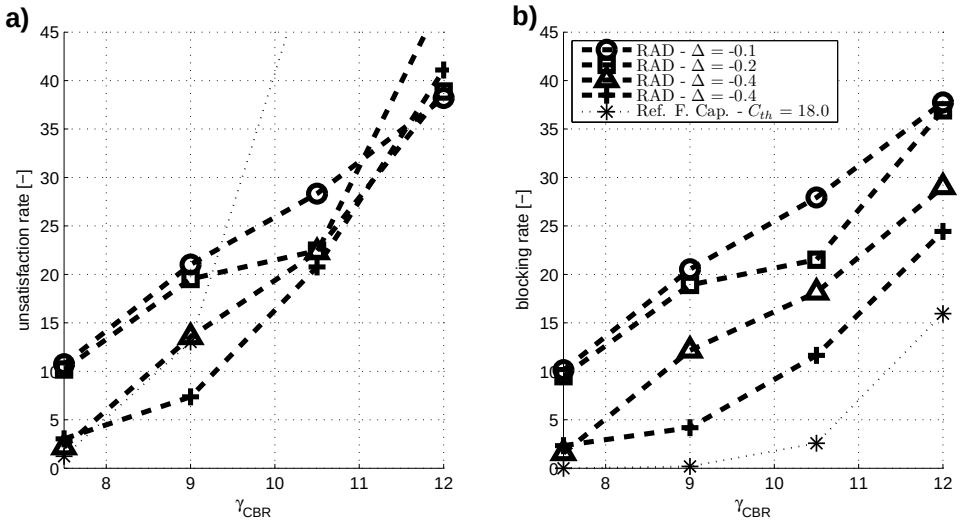

Figure 6.8: Overall performance of the fixed RAD based RAC algorithm. Figure (a) shows the unsatisfaction rate and Figure (b) shows the blocking rate

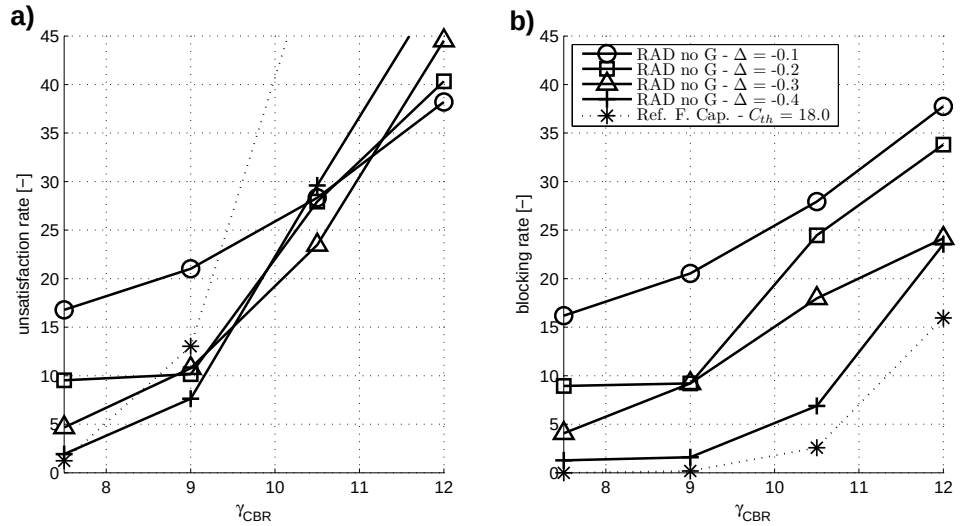

Figure 6.9: Overall performance of the fixed RAD based RAC algorithm without taking into account the Geometry information. Figure (a) shows the unsatisfaction rate and Figure (b) shows the blocking rate

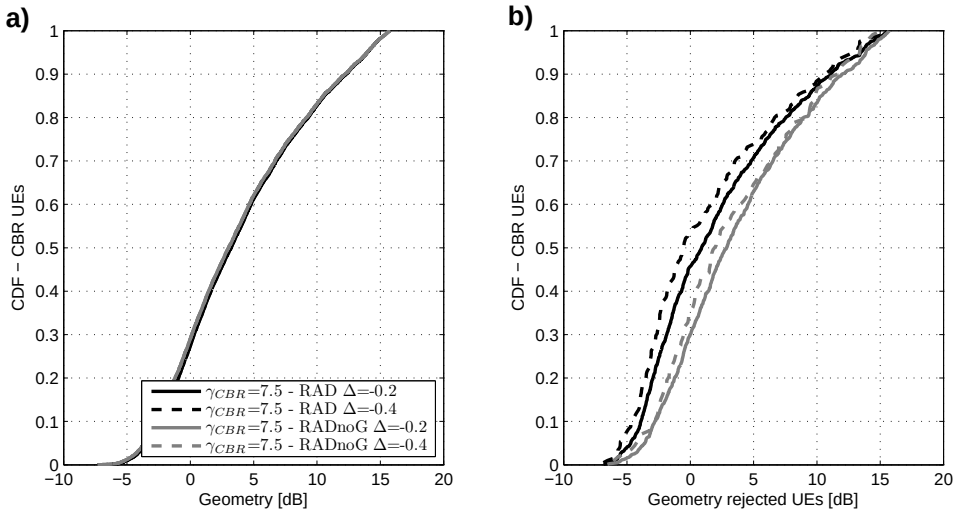

Figure 6.10: Influence of the geometry estimation on the RAD based RAC for a CBR offered load of $\overline{\gamma_{CBR}} = 7.5$Mbps. (a) shows the geometry CDF of the UEs in the system and (b) shows the geometry CDF of the UEs blocked by the RAC. Both results for RAD-based RAC taking into account the geometry and not taking into account the geometry are presented.

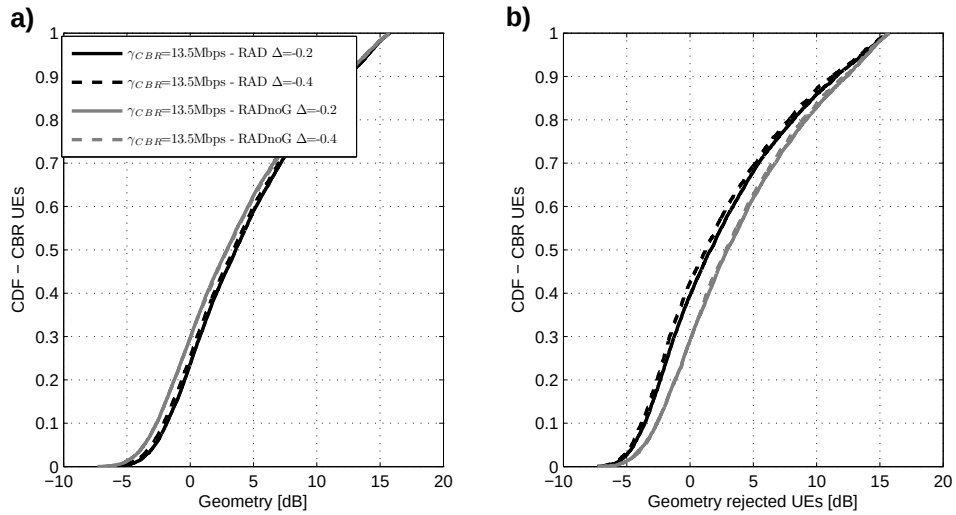

Figure 6.11: Influence of the geometry estimation on the RAD based RAC for a CBR offered load of $\overline{\gamma_{CBR}} = 7.5$Mbps. (a) shows the geometry CDF of the UEs in the system and (b) shows the geometry CDF of the UEs blocked by the RAC. Both results for RAD-based RAC taking into account the geometry and not taking into account the geometry are presented.

6.5.6.1 The Influence of the geometry Estimation on the RAD based RAC

Figures 6.10 and 6.11 represent the geometry statistics for two different offered loads: $\overline{\gamma_{CBR}}$ = 7.5Mbps and $\overline{\gamma_{CBR}}$ = 12Mbps. (a) shows the geometry distribution of UEs accepted into the system and (b) shows the geometry distribution of blocked UEs. For the two different cases, the version of RAD-based RAC that takes into account the geometry information tends to block more low geometry UEs. However, the impact on the geometry distribution of accepted UEs can be seen only for the case with $\overline{\gamma_{CBR}}$ = 12Mbps. The reason is that at $\overline{\gamma_{CBR}}$ = 7.5Mbps, the blocking rate is low (according to Figures 6.10 and 6.11, the blocking rate is situated between 2.5% and 10%) while, for $\overline{\gamma_{CBR}}$ = 12Mbps, the blocking rate is much higher (between 25% and 35%). Therefore, the effect of taking into account the geometry information on the geometry distribution of accepted UEs obviously increases with the offered load. Another aspect which is important to consider is the precision of the geometry information. Indeed, in the model described in Section 6.5.1, the geometry:

- is subject to measurement errors
- cannot be evaluated above 0dB

Those two facts contribute to a lack of precision of the geometry information. This lack of precision is also a factor limiting the capacity of the RAD-based RAC to differentiate UEs according to their geometry and clearly limits the impact of geometry based required activity of incoming UEs estimation.

6.5.6.2 Channel trackability of the RAD based RAC algorithm

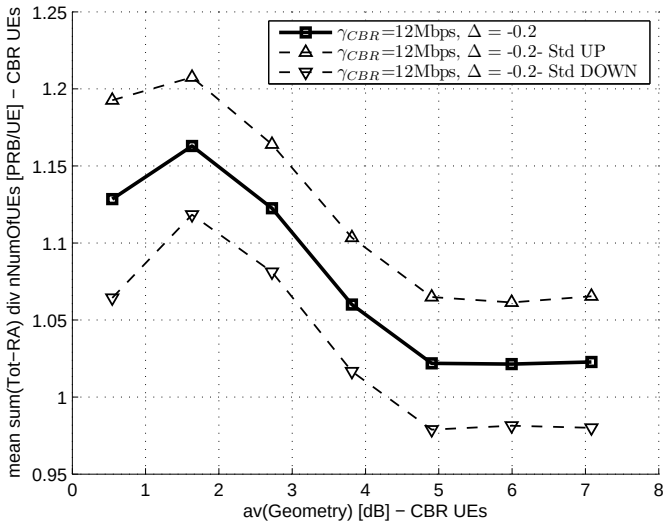

Figure 6.12: average geometry versus mean ratio between sum of required activities divided by the number of CBR UEs.

Figure 6.12 shows the mean geometry ($\frac{1}{N_{UE}} \cdot \sum_{n=1}^{N_{UE}} \mathbf{G_{dB}}\,[n]$) versus the corresponding average values of the mean required activity ($\frac{1}{N_{UE}} \cdot \sum_{n=1}^{N_{UE}} \mathbf{RA}\,[n]$). The plot is built from values sampled every TTI. The standard deviation of the mean required activity is expressed on the Figure

with additional points above and below the main curve. The mean geometry represents the general channel quality of the cell. If more UEs are close to the eNode-B, the mean geometry will increase and if UEs get further form the eNode-B, then the mean geometry will decrease. The main interest of the mean geometry is that it is an objective measurement that is made from physical datas. On the contrary, the mean required activity is a subjective measure in the sense that it is built by the data accessible by the eNode-B. The common ground between those two measures is that they both aim at evaluating the average channel condition in a cell. Comparing those two measures can therefore help judging to a certain extent the accuracy of the required activity detection process.

Figure 6.12 shows that the mean required activity globally decreases while the mean geometry increases. Indeed, when the global channel quality increases, it means that globally, the UEs have a better channel quality and therefore need less scheduling time to achieve their GBR. The Figure therefore proves that the required activity detection process is consistent with the UEs channel quality. This is also the proof that the RAD RAC algorithm is able to track the average channel condition of the cell.

However, the standard deviation of the mean required activity indicates that the precision of the required activity is limited as it is fairly large thought not as large as the total span of the mean required activity. This limited precision is mainly due to:

- limited and error prone CQI feedback
- the estimation of $\mathbf{Rsch}$ is ever changing and is subject to large errors at initialization.

Therefore, RAD RAC can track the channel but with a limited precision due the lack of accuracy of the information available at the eNode-B. This is confirmed by the results in Figure 6.2 showing that the RAD-based RAC with no Geometry information does not outperform in terms of UE satisfaction the fixed capacity algorithm (with C_{th} = 13.3Mbps).

It is important to note here that both statistics: mean geometry and mean required activity are based on CBR UEs exclusively. It had been seen for the adaptive capacity RAC, that finite buffer UEs staying for too long in the network create a bias for the capacity estimation. It cannot be the case for the required activity estimation, since it is based only on CBR UEs statistics and therefore the correlation between the mean geometry of CBR UEs and mean required activity is not affected by finite buffer UEs in any way. In that sense, the RAD-based RAC is superior to the adaptive capacity based RAC.

6.5.6.3 Conclusion on the RAD-based RAC

Finally, the RAD-based RAC is an interesting algorithm for two main reasons:

- It is able to track the average channel quality within certain limits. Therefore, it requires only limited tuning while the fixed capacity algorithm would required tailored tuning for each new cell.
- Taking the geometry information into account provides extra UE satisfaction gain. Figure 6.2 shows that taking the geometry information into account reduces the unsatisfaction rate of 33% for $\overline{\gamma_{CBR}} = 10.5$Mbps.

Finally, the RAD based RAC with geometry information is the best RAC algorithm presented in this study since it provides the smallest unsatisfaction rate. It provides an unsatisfaction loss of 12% compared to the fixed throughput based RAC ($C_{th} = 13.3$Mbps) algorithm at $\overline{\gamma_{CBR}} = 10.5$Mbps.

6.6 Conclusion

In this Chapter, several RAC are derived for the LTE framework developed in the present thesis. Algorithms from the literature like fixed capacity and adaptive capacity based RAC are thoroughly details and explained. A new algorithm based on the RAD principle is also introduced. The different algorithms are tested in the LTE framework developed along the thesis for a mixed traffic case with different traffic offered load values. The RAD algorithm is tested with and without taking into account geometry information of incoming UEs.

The fixed throughput based RAC algorithm provides very good results over a reference case with almost systematic UE admission. The fixed throughput based RAC can reduce the UE unsatisfaction rate of up to 30% for $\overline{\gamma_{CBR}} = 9$Mbps and up to 50% for $\overline{\gamma_{CBR}} = 10.5$Mbps compared to the reference case. However, the fixed throughput based algorithm need to be tuned for every cell. Therefore, auto tunable algorithms are considered: the adaptive throughput algorithms and the RAD based RAC algorithm.

It is concluded that the adaptive throughput based RAC algorithm is not suitable in the present form as the best effort UEs bias the capacity estimate. It could however be improved by not taking into account the best effort UEs and building the estimate based only on GBR UEs. On the other hand the RAD-based RAC can reduce further the unsatisfaction rate provided by the fixed capacity algorithm. Indeed, over the best tuning of the fixed capacity algorithm, the unsatisfaction rate is reduced of 12% for $\overline{\gamma_{CBR}} = 10.5$Mbps. Finally, The RAD-based RAC is recommanded for implementation in a real Orthogonal Frequency Division Multiple Access (OFDMA) downlink system since it does not require any cell specific tuning and since it gives the best performance.

Chapter 7

Conclusion

In this thesis, Quality of Service (QoS) aware Radio Resource Management (RRM) solutions are proposed for downlink Orthogonal Frequency Division Multiple Access (OFDMA) wireless mobile access systems. The study focuses on QoS aware Radio Admission Control (RAC) and Packet Scheduler (PS). The study aims at providing algorithms that are applicable in a real system and therefore low complexity and noisy channel feedback are included in the design constraints. The 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) is taken as a case study. The different algorithms are tested in a multicell system level simulator following the LTE standard. Functionalities like Link Adaptation (LA), Outer Loop Link Adaptation (OLLA) and Hybrid Automatic Repeat reQuest (HARQ) are explicitly implemented. The Channel Quality Information (CQI) reporting scheme takes into account uplink signaling channel limitations and includes error modeling.

7.1 Recommendations for QoS aware Packet Scheduler

The decoupled packet scheduling design in Time Domain Packet Scheduling (TDPS) and Frequency-Domain Packet Scheduling (FDPS) aims at offering a low complex packet scheduler, independently of the number of users to schedule. Throughput control has not been studied widely in the literature for this type of decoupled packet scheduler. In the thesis, two principles regarding throughput control recommended to observe when designing a decoupled packet scheduler are highlighted. Firstly, the time and frequency domain schedulers should be independent in the sense that they should not include contradictory throughput control mechanisms. Secondly, the throughput control should be managed by the TDPS if the number of users is large. In this case, the FDPS is left with the role of maximizing the cell throughput by taking advantage of the multi user diversity. The throughput control should however be managed by the FDPS when the number of user is low. The metric weighting can be used for that purpose. Furthermore, two FDPS algorithms including no throughput control mechanism are introduced: Carrier over Interference to Average and Proportional Fair scheduled. Those packet schedulers are recommended to be implemented in a real system as they provide a cell throughput gain of the order of 10% compared to the reference FDPS Throughput To Average. Then a full Quality of Service aware packet scheduler for traffic mix (best effort and real time traffics) is developed based on the Required Activity Detection (RAD) principle. The Modified Largest Weighted Delay First (M-LWDF) time domain scheduler is recommended as time domain scheduler and preferred to the RAD time domain sched-

uler such as derived in the thesis. However, based on the detailed analysis of the algorithms, it is believed that RAD time domain scheduler can be further improved. Finally, the RAD frequency domain scheduler is recommended for implementation as a complement to Carrier over Interference to Average (CoItA) or Proportional Fair scheduled (PFsch) as it brings a outage decrease of up to 50% in the simulated cases.

7.2 Recommendations for handling fractional load situations

It is shown that when a fractional load situation occurs, the BLock Error Rate (BLER) can increase up to a value close to 100% if a simple packet scheduler is applied to the packet scheduler. Such an increase is non disirable in terms of QoS provision as it increases the packet delivery delays. Therefore several algorithms are introduced in order to overcome this BLER increase.

Physical Resource Block (PRB) pattern selection are introduced and prove to efficiently solve the BLER increase. The best PRB pattern selection algorithm called Best Quality PRB Pattern (BQPP) can even increase the spectral efficiency of up to 120% compared to the full load case since it opportunistically choses to transmit on the PRBs with the lowest interference. Another solution is the application of Wideband Interference Reporting (WIR) in the Channel Quality Information reporting scheme as it also solves the BLER increase problem. Even if WIR does not provide a significantly lower specral efficiency gain than BQPP, it is the prefered and recommanded solution as it is the simplest and extra capacity is not needed when the offered traffic is not bandwidth limited.

Furthermore, it is shown that the algorithms presented in the thesis cannot be used as an Inter-Cell Interference Coordination (ICIC) technique.

7.3 Recommendations for QoS aware Radio Admission Control

Radio Admission Control (RAC) is necessary to avoid the system to be overloaded and to give the possibility to the packet scheduler to handle the offered traffic.

It is shown that a simple Radio Admission Control (RAC) scheme called fixed throughput RAC can perform User Equipment (UE) blocking and keep a cell in its feasibility region. However, this algorithms needs to be fine tuned for every different type of cell. Therefore, two adaptive algorithms are also proposed and studied: the adaptive throughput based and RAD-based RAC. The RAD-based RAC is the best tested algorithm and can bring UE satisfaction gain over the fixed capacity algorithms only if the Geometry information of the UE is taken into account. In that case, the unsatisfaction rate is reduced of 12% for a Constant Bit Rate offered load of 10.5Mbps compared to the best tunning of the fixed throughput algorithm. The RAD-based RAC is therefore recommended for implementation in a real system.

7.4 Future Works

Firstly improvement can be made on the design of the presented algorithm. Specifically the RAD concept can be further developed as a TDPS algorithms and FDPS algorithms. Suggestions in that

direction have been made along the thesis. The RAD based RAC algorithm can also be further improved.

Besides, a next step in the study of RRM for OFDMA donwlink system is a detailed study of Voice Over Internet Protocol (VoIP) traffic. Indeed, VoIP will be one of the major service used in future Orthogonal Frequency Division Multiplexing (OFDM) network and it is a very challenging traffic due to its low throughput and the connected signaling constraints.

Appendix A

Simulation methodology: Simulator optimization

Fast Fading Implementation Optimization in an OFDMA System Simulator

G. Monghal*, I. Z. Kovács†, A. Pokhariyal*, K. I. Pedersen†, C. Rosa†, P. E. Mogensen*†
*Aalborg University, †Nokia Networks
9220 Aalborg Ø, Denmark
Email: gm@kom.aau.dk

***Abstract*— In this paper, we investigate methods for improving an OFDMA system simulator in terms of trade-off between complexity and radio channel implementation accuracy. Fast fading and SINR calculations are the most resource consuming tasks of a system simulator. In order to decrease their complexity, it is proposed to reduce the number of calculated Channel Transfer Function (CTF) realizations according to the time and frequency correlation properties of the simulated radio channel. Besides, we highlight that the upsampling of the Channel Impulse Response (CIR), which is needed for the Fast Fourier Transform (FFT) application, may introduce major errors in the channel frequency statistics if it is done by simple Brute Force Grid Alignment (BFGA) method. To achieve better results we introduce a low complex upsampling method which reduces the simulation inaccuracies significantly. The possibility to trade complexity for precision is shown. It is possible for instance to divide the Fast Fading and SINR calculation complexity by 50 while keeping a reasonable simulation accuracy.**

I. Introduction

OFDMA (Orthogonal Frequency Division Multiple Access) has been chosen in the latest Wireless cellular network standards. It has been included in the IEEE 802.16 [0] (also known as WiMAX) and later in the 3GPP UTRAN LTE (Long Term Evolution) [0]. For this reason, research is very active to provide general system performance results for the different standards. Generally, system performance is investigated and evaluated through a multicell system level simulator. Such simulators are subject to high complexity for two reasons. Firstly they must reproduce natural fading conditions of a great number of radio links, especially in case of Multiple Input Multiple Output (MIMO) transmissions [0]. Secondly, the Signal to Interference and Noise Ratio (SINR) calculation is made in the frequency domain on a subcarrier basis. Therefore, simulation speed is obviously a critical issue for research productivity.

This paper studies speed improvement by proposing a method which reduces the fast fading generation time. This method consists of reducing the resolution of calculated Channel Transfer Function (CTF) values in both time and frequency. Firstly, a proper resolution is determined with the time and frequency correlation properties of the simulated channel. Depending on this resolution, a minimal FFT size is chosen together with a Channel Impulse Response (CIR) up-sampling rate. The CTF can be calculated by simply applying the FFT on a vector, upsampled with Brute Force Grid Alignment BFGA from the CIR. However, this may in given conditions cause a major difference in the channel statistics. In order to overcome this problem, we present a CTF calculation method based on the Minimum Square Error (MSE) criterion where the FFT is applied on a filtered CIR.

The rest of the paper is organized as follows: The basic principles of an OFDMA system simulator including channel modeling are outlined in Section II. A strategy for choosing a resolution in time and frequency based on the correlation properties is introduced in Section III. Section IV describes the minimal FFT size and CIR upsampling rate selection as well as the different CTF calculation methods. Validation and speed improvement results are presented in Section V. Finally, the paper is closed with conclusions in Section VI.

II. OFDMA Downlink System Simulator

A. System Modeling

A cellular system can be modeled by a number of M_s Mobile Stations (MS) and B_s Base Stations (BS). Each radio link between MS and BS can be either a desired signal link or an interfering link: In total, the system comprises $B_s \times M_s$ radio links. On each of those links, the fading process has to be simulated independently. Furthermore, if a MIMO scheme is simulated with N_{Rx} receive antennas and N_{Tx} transmit antennas, then $N_{Rx} \times N_{Tx}$ channels are simulated per link, meaning, $N_{Rx} \times N_{Tx} \times B_s \times M_s$ links in total.

In OFDMA, the bandwidth is divided into N_s subcarriers which are Δf_s-spaced. Subcarriers are grouped into N_{RB} Resource Blocks (RB) being the scheduling unit. A popular simulation method for estimating the packet arrival performance is the Effective SINR [0] approach. In this method, the SINR is calculated on a subcarrier basis. Then, the subcarrier SINRs of one block are mapped into one effective SINR value. The block error rate is determined from this value through correspondance tables built from link level simulations.

B. Channel Modeling

A link between a BS and a MS is determined by its path loss, shadow fading and fast fading. The fast fading describes the small scale channel variations caused by multipath and the movements of the MS. For the sake of simplicity, we consider the fast fading effects to be constant over one OFDM sub-frame of length Δt_s denoted by index n. A widely used channel model is the Wide Sense Stationary Uncorrelated Scatterers WSSUS) model with Clarke's Doppler Spectra [0].

In this model, the CIR expresses the state of the channel at time step n:

$$c_n(\tau) = \sum_{p=0}^{N_p-1} \mathbf{c}_n[p]\delta(\tau - \tau_p) \quad (1)$$

The sequences $(\mathbf{c}_n[p])_n$ are complex Gaussian fading processes with average power equal to P_p. The two sets of data P_p and τ_p define the Power Delay Profile (PDP) of the channel where τ_p are the tap delays and P_p the paths powers. N_p is the number of path in the PDP.

In OFDM, the calculation of the SINR experienced on each subcarrier, is based on the CTF. The CTF on the subcarrier central frequencies is calculated by Fourier transform. The expression of the CTF on the k^{th} subcarrier is:

$$\mathbf{C}_n[k] = \sum_{p=0}^{N_p-1} \mathbf{c}_n[p]e^{-2i\pi\tau_p k\Delta f_s} \quad (2)$$

We point out that the expression of the time domain autocorrelation in Clarke's model is expressed by a zero-th order Bessel function of the first kind [0] [0]:

$$\phi_{\mathbf{C}}^{time}(\Delta t) = J_0(2\pi\frac{vf_c}{c}\Delta t) \quad (3)$$

where c is the speed of light, f_c the carrier frequency and v is the velocity of the MS. Similarly, the frequency correlation of the CTF is defined by the fourier transform of the PDP:

$$\phi_{\mathbf{C}}^{freq}(\Delta f) = \sum_{p=0}^{N_p} P_p e^{-2i\pi\tau_p\Delta f} \quad (4)$$

III. Time and Frequency Resolution Optimization

The principle introduced in this section consists in optimizing the speed of the fast fading generation by reducing the resolution of the CTF in both time and frequency domains. This principle is illustrated in Figure . The transfer function can be seen as a grid of values along two dimensions namely time and frequency. The grid units are:

- The subcarrier spacing in the frequency dimension, Δf_s.
- The sub-frame in the time dimension, Δt_s.

In the following we focus on calculating the two following parameters:

- M, the time resolution of the CTF.
- K, the frequency resolution.

We call this process "resolution reduction". Parameters M and K need to be chosen carefully since they should be small enough not to introduce bias in the simulation results. We base the choice of these parameters on the mean square error of the CTF defined as:

$$E(\epsilon_{K,M}^2) = E(|\mathbf{C}_m[k] - \mathbf{C}_{m+M}[k+K]|^2)_{k,m} \quad (5)$$

Analysing the expression of the mean square error, we see that time and frequency correlation have the same influence on the error when a resolution reduction is performed.

$$E(\epsilon_{K,M}^2) = 2(1 - Re(\phi_{\mathbf{C}}^{freq}(K\Delta f_s)\phi_{\mathbf{C}}^{time}(M\Delta t_s))) \quad (6)$$

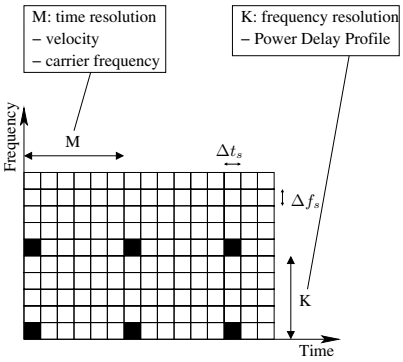

Fig. 1. Frequency and time resolution of the Channel Transfer Function.

Preferably, the choice of K and M should be so that the mean square error is distributed equally between time and frequency domains, with the choice of K depending on the PDP in (4) and the choice of M depending on the MS velocity and carrier frequency in (3). The mean square error in (6) can be used as a measure for the error we commit in simulation by resolution reduction.

IV. Channel Impulse Response Upsampling and FFT

The CTF can be calculated with a Fast Fourier Transform (FFT) applied on the upsampled CIR. This section describes firstly an algorithm for choosing the two parameters: optimal FFT size N_{FFT} and upsampling time $\Delta\tau$ according to the frequency resolution K. Secondly, we introduce two different upsampling methods: The Brute-Force Grid Alignment (BFGA) and a new method called Minimum Square Error (MSE) upsampling.

A. FFT Parameters

In order to calculate the CTF, the CIR must be upsampled with a sampling time $\Delta\tau$ into a vector of size N_{FFT}. For the resolution $K = 1$, those two parameters can be taken from the OFDM transceiver parameters. Here we introduce a general algorithm which calculates the optimal FFT size for any given frequency resolution:

- The FFT order N_{FFT} is chosen as the smallest power of two greater than $\frac{N_s}{K}$:

$$N_{FFT} = 2^{\lceil \log_2 \lceil \frac{N_s}{K} \rceil \rceil} \quad (7)$$

- The sampling time $\Delta\tau$ must satisfy a subcarrier spacing of K at the output of the FFT. This condition is respected with:

$$\Delta\tau = \frac{1}{N_{FFT}K\Delta f_s} \quad (8)$$

- However, the upsampled CIR must be long enough to fit the maximum excess delay of the CIR. This introduces a

condition on K:

$$K < \frac{1}{\tau_{N_p-1}\Delta f_s} \quad (9)$$

- Finally, the CTF should have at least one value calculated per RB:

$$K \leq \frac{N_s}{N_{RB}} \quad (10)$$

In Table different values of N_{FFT} and $\Delta\tau$ are given for two different LTE configurations: $N_s = 300$ and $N_s = 600$ (while $\Delta f_s = 15kHz$).

TABLE I
UPSAMPLING PARAMETERS FOR LTE PARAMETERS: $\Delta f_s = 15kHz$ AND $N_s = 300$ - $Ns = 600$.

	$N_s = 300$ $BdW = 4.5MHz$		$N_s = 600$ $BdW = 9MHz$	
K	N_{FFT}	$\Delta\tau[ns]$	N_{FFT}	$\Delta\tau[ns]$
1	512	130.21	1024	65.104
2	256	130.21	512	65.104
3	128	173.61	256	86.806
4	128	130.21	256	65.104
5	64	208.34	128	104.17
8	64	130.21	128	65.104
10	32	208.34	64	104.17
16	32	130.21	64	65.104
25	16	166.66	32	83.333

B. Channel Impulse Response Upsampling

1) Brute-Force Grid Alignment: The principle of BFGA is illustrated in Figure . In this method, the FFT is applied on a zero-padded vector $\mathbf{a}_n$ of size N_{FFT} where $\mathbf{a}_n[q]$ is the sum of the elements $\mathbf{c}_n[p]$ which verify:

$$q = round(\frac{\tau_p}{\Delta\tau}) \quad (11)$$

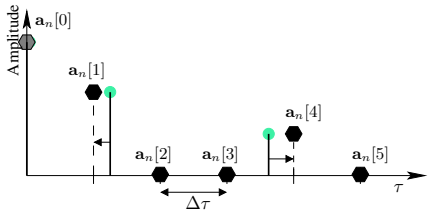

Fig. 2. Brute Force Grid Alignment illustration with a 3 tap CIR.

However, this method may be very inaccurate for actual representation of the channel. Indeed, in some cases, the sampling time $\Delta\tau$ may be significantly larger than the tap delay differences in the original PDP. For example, the tap delays 2, 3 and 4 of the PDP Typical Urban [0] are separated by only $5ns$ whereas the minimum value of $\Delta\tau$ in Table is $65.104ns$.

2) Minimum Square Error: Here we describe a new upsampling approach which consists in calculating the vector $\mathbf{b}_n$ to minimize the square distance between $FFT(\mathbf{b}_n)$ and the actual transfer function. In order to find a general solution for $\mathbf{b}_n$, we use an approach based on vectorial space. The FFT produces a linear combination of discretized functions. Those functions consist of an orthonormal base of a subvectorial space S of $L^2(0, N_s\Delta f_s)$ where $L^2(0, N_s\Delta f_s)$ is the space of the continuous, square integrable functions on $[0, N_s\Delta f_s]$. Therefore, $FFT(\mathbf{b}_n)$ is the orthogonal projection of the transfer function on S. By deriving simple scalar products between the transfer function and the FFT base functions, we can easily express $\mathbf{b}_n$ as the product $\mathbf{\Upsilon}\mathbf{c}_n$ where:

$$\mathbf{\Upsilon}[l,p] = \begin{cases} 1 & \text{if} \quad l\Delta\tau = \tau_p \\ \frac{e^{2i\pi(l-\frac{\tau_p}{\Delta\tau})}-1}{2i\pi(l-\frac{\tau_p}{\Delta\tau})} & \text{otherwise} \end{cases} \quad (12)$$

In practice, this upsampling method consists of spreading the power of the original taps over the vector on which the FFT is applied. Notice that in each column, the matrix $\mathbf{\Upsilon}$ concentrates high values around $l = \frac{\tau_p}{\Delta\tau}$. Therefore, in order to decrease the complexity of this upsampling process, we propose to select in each column of $\mathbf{\Upsilon}$ only the larger values by applying an N_{MSE} size window. This keeps most of the information contained inside the matrix and may reduce the complexity significantly. Figure gives an illustration of the power spread over the vector and of the windowing.

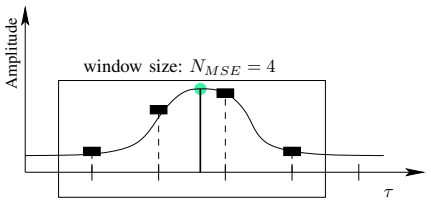

Fig. 3. Minimum Square Error upsampling: illustration of the tap power spreading along the vector $\mathbf{b}_n$ for a window size of $N_{MSE} = 4$.

V. RESULTS

The following part shows the impact of the different upsampling methods on the frequency statistics of the channel. Then, possible trade-offs between complexity and simulation precision are presented.

A. Frequency Correlation Properties

As mentioned in the previous section, upsampling of the CIR with BFGA may have significant effects on the frequency channel statistics. We give here several examples of those effects on different Power Delay Profiles. We also show the improvement brought by MSE upsampling with $N_{MSE} = 4$.

In Figure we can see that the statistics resulting from BFGA upsampling with a sampling time of $\Delta\tau = 104.17$ns do not change significantly the frequency correlation of the channel. Therefore, it is acceptable to perform system simulation with these settings although we can see that by applying MSE upsampling, the generated channel correlation properties fits almost perfectly to the theoretical curve.

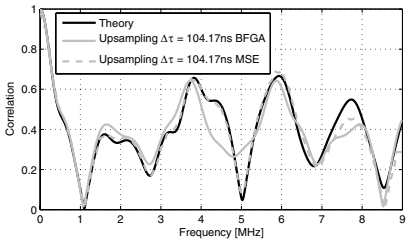

Fig. 4. Typical Urban frequency correlation: effects of upsamplings BFGA and MSE with N_{MSE} = 4 and $K = 10$.

Figure highlights a very different case where BFGA upsampling changes significantly the channel statistics of PDP ITU Indoor A [0]. On the contrary, MSE upsampling generates channel statistics which are very close to the theoretical statistics up to a certain frequency offset. Indeed, The error is lower bounded by the average distance between the subspace S and the CIR. We also observe that with BFGA the frequency correlation varies with different $\Delta\tau$. For example $\Delta\tau = 65.104$ns provides a better fit to the thoretical curve than $\Delta\tau = 104.17$ns. Nevertheless, simulating the channel with BFGA may require manual test and tuning prior to performing simulations to ensure the simulation accuracy. It also brings limitations in the choice of the frequency resolution K. However, MSE upsampling minimizes error in the simulation of the frequency channel statistics.

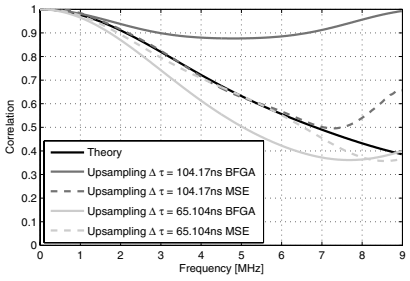

Fig. 5. Indoor A frequency correlation: effects of upsamplings BFGA and MSE with N_{MSE} = 4.

B. Resolution Reduction Errors

In order to evaluate the error produced when reducing the resolution, we use an OFDMA system model with characteristics summarized in Table . This scenario corresponds to LTE Macro Case 1 described in [0]. In this layout the 3 central sectors are the simulated sectors in which MSs are dropped for a fixed amount of time. All MS locations in the serving sectors have equal probability. The other cells are interfering cells. The

TABLE II
OFDMA SYSTEM AND CHANNEL PARAMETERS.

System Parameter	Setting
Cellular Layout	19 sites, 3 sectors per sites
Inter-site distance	500m
Total BS transmit power	46 dBm
Penetration loss	20 dBm
Shadowing Standard deviation	8 dB
Subcarrier Spacing, Δf_s	15 kHz
Number of Data Subcarriers, N_s	600
Number of RB, N_{RB}	24
Sub-frame duration, Δt_s	0.5ms
MS Velocity, v	3kmph
Carrier Frequency, f_c	2GHz
PDP	Typical Urban 20 taps [0]

validity estimation of Resolution Reduction is evaluated with two metrics. Firstly, the SINR (in dB) error standard deviation

$$\Delta SINR_{std}^{K,M} = Std\big[SINR_{k,l,n}^{K,M} - SINR_{k,l,n}^{1,1}\big]_{k,l,n} \quad (13)$$

where $SINR_{k,l,n}^{K,M}$ is the SINR on subcarrier k, for the MS l at sub-frame n. $SINR_{std}$ characterizes the error introduced by Resolution Reduction on the subcarrier SINR. We also use a second metric which describes the error by resolution reduction in the effective SINR. The chosen effective SINR modeling is Exponential Effective SINR $\Delta SINRe$. As block size we take the smallest system unit: The RB in one sub-frame.

$$\Delta SINRe_{std}^{K,M} = Std\big[SINRe_{q,l,n}^{K,M} - SINRe_{q,l,n}^{1,1}\big]_{q,l,n} \quad (14)$$

where $SINRe_{q,l,n}^{K,M}$ is the SINR on RB q, for the MS l and at sub-frame n. All simulations have been run with the same sampling time ($\Delta\tau = 65.104$ns) in order to give relevant metric estimations (see table).

Figure shows the two error metrics with Frequency Domain Resolution Reduction only ($M = 1$). Our first conclusion is that both SINR and effective SINR errors increase with the resolution K but we observe that the Effective SINR error increases less than the SINR. This is due to the fact that the effective SINR is an (exponential) averaging of the SINR values within a RB. The averaging process divides the error as well, as long as the calculated values are sufficiently correlated.

Figure shows the two error metrics with time Domain Resolution Reduction ($K = 1$) only. Here contrary to Frequency Domain Resolution Reduction, Effective SINR and

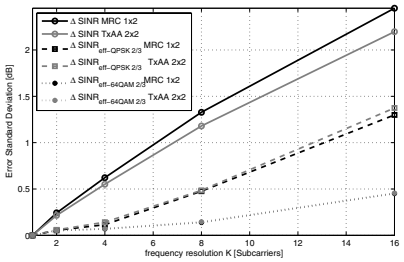

Fig. 6. Effects of frequency domain resolution reduction with M=1.

SINR have the same magnitude of error. Indeed, in our model no averaging reduces the error, since the effective SINR is calculated every sub-frame. Furthermore, we simulated two different multiple antenna transmission schemes:

- Maximum Ratio Combining (MRC) with 1 transmit antenna and 2 receive antennas (1x2)
- Transmit Adaptibe Array (TxAA) with 2 transmit antennas and 2 receive antenna (2x2)

We observe in both figure and that the error is independant of the transmission scheme.

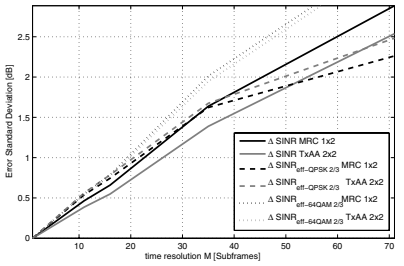

Fig. 7. Effects of time domain resolution reduction with K=1.

Figure shows an example of the tradeoff for Frequency and Time domain Resolution Reduction. It is possible to keep a reasonably low error of 0.5dB in average for $K = 4/M = 11$.

C. Complexity Gain

Figure shows the complexity gain $R(K, M)$ of the fast fading and SINR calculation process compared to a configuration with resolution $K = 1$ and $M = 1$ where $N_{FFT}(K)$ is calculated depending on K with the algorithm described in section IV.

$$R(K, M) = \frac{1}{M} \frac{N_{FFT}(K) \log 2(N_{FFT}(K))}{N_{FFT}(1) \log 2(N_{FFT}(1))} \quad (15)$$

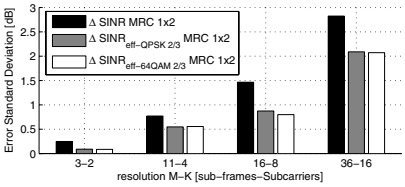

Fig. 8. Time and frequency domain resolution reduction with different time and frequency resolutions.

This shows the potential system simulator speed improvement. We see that with resolution $K = 4/M = 11$, the processing time of fast fading and SINR calculation is divided by 50. Fig. and show the tradeoffs between complexity and precision possible to achieve when simulating an LTE configuration with 10MHz system bandwidth.

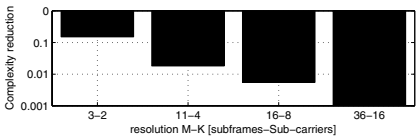

Fig. 9. Complexity gain in Fast Fading and SINR calculation for different time and frequency resolution relatively to resolution K=1 - M=1.

VI. Conclusion

We have investigated methods for optimizing the speed and controling errors due to channel modeling in an OFDMA system level simulator. With the different methods we introduced, it is possible to achieve tradeoffs between complexity and simulation precision. For example, we investigated an LTE configuration with a 10MHz system bandwidth simulated with Typical Urban PDP. We showed that it is possible to reduce the complexity of fast fading generation and SINR calculation by a factor of 50 while keeping the average interpolation error of the Effective SINR at 0.5 dB. Furthermore, we introduced a CIR upsampling method which minimizes the MSE.

References

[1] 3GPP Technical Report 25.814 Version 7.0.0. *Physical Layer Aspects for Evolved UTRA*, June 2006.
[2] IEEE std 802.16e-2005 *Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and MobileOperation in Licensed bands*, february 2006.
[3] K. Brueninghaus, D. Astley, T. Salzer, S. Visuri, A. Alexiou, S. Karger and G.-A. Seragi, "Link Performance Model for System Level Simulations of Broadband Radio Access Systems", in *Proceedings of IEEE Personal Indoor and Mobile Radio Communication (PIMRC)* vol. 4, Berlin, Germany, September 2005, pp. 2306-2311.
[4] W. C. Jakes "Microwave Mobile Communication", 1974
[5] J. D. Parsons "The Mobile Radio Propagation Channel, Second Edition", 2000
[6] 3GPP Technical Report 101 112 version 3.2.0 *Selection Procedure for the Choice of Radio Transmission technologies of the UMTS*, april 1998

Appendix B

Simulations Statistical Relevance

B.1 Introduction

This appendix describes The general method to assess the statistical relevance of most of the simulations in the thesis. Representative examples of the results presented in the thesis are shown. This appendix deals specifically with the statistical relevance of the average cell throughput and coverage as those two performance indicator are widely used along the thesis. Firstly, the assessment method for infinite buffer simulations is presented, then for finite buffer.

B.2 Infinite Buffer

B.2.1 Average cell throughput

For the infinite buffer simulation modeling, the following probability space is constructed:

$$PR_{N_{UE}} = (\Omega, \mathcal{F}, P) \tag{B.1}$$

where the sample space Ω is the set of possible combinations of N_{UE} cell physical position and shadow fading value:

$$\begin{aligned}\Omega = \{\, &(x, y, S)^{N_{UE}} \,/ \\ &(x, y) \text{ are the possible position coordinates and} \\ &S \text{ are the possible shadowing values}\}\end{aligned}$$

$\mathcal{F}$ is set of subsets of Ω and P is a measure on Ω that corresponds to the probability of each element of Ω (it is not the point in the appendix to describe fully the probability space but to use it as a tool to understand the principle of infinite buffer simulation and how to asses the statistical relevance).

The Cell throughput is defined as a random variable X_{PS} on $PR_{N_{UE}}$ that depends on the PS algorithm PS:

$$\begin{aligned} X_{PS}: \quad &\Omega \to \mathbb{R} \\ &\omega \to x \end{aligned} \tag{B.2}$$

and the average cell throughput is defined as the expectation of X_{PS}: $\mu_{PS} = E(X_{PS})$.

ω corresponds to an instantiation of Ω, it corresponds also to a "run" of an infinite buffer simulation. The value x corresponds to the converged mean cell throughput (mean cell throughput of all the cell throughput per Transmission Time Interval (TTI)) in one "run". Note that therefore, the definition of X_{PS} implies that the mean cell throughput converges. In practice, in a simulation as the run length cannot be infinite, it is impossible to have the mean cell throughput converging, however the run length should be long enough to provide a minimal error.

An infinite buffer simulation can be modeled as a set of Independent and Identically Distributed (i.i.d) random variables X_{PS}^i with $i \in [1, N]$ and where N is the number of runs. The average cell throughput in the thesis is estimated with the sample mean estimator:

$$\hat{\mu}_{PS}^N = \sum_{i=1}^{N} \frac{X_{PS}^i}{N} \tag{B.3}$$

The relevance of the estimator $\hat{\mu}_{PS}^N$ can be evaluated thanks to its standard deviation estimate:

$$\hat{\sigma}_{PS}^N = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{(N-1)} \sum_{i=1}^{N} \left(X_{PS}^i - \hat{\mu}_{PS}^N\right)^2} \tag{B.4}$$

It is chosen to show a characteristic interval of $s = 2 \cdot \hat{\sigma}_{PS}^N$ above and below the sample mean in order to evaluate the precision of the infinite buffer simulations cell average throughput estimate.

Table B.1: Infinite buffer simulation for evaluation of statistical significance

Simulation	Run length	Num of runs	N_{UE}	Section
BL RR	10s	150	30	2.6.5
BL PF	10s	150	30	2.6.5
CoItA/PSS 400kbps	10s	150	30	3.7.1
CoItA/PSS 1100kbps	10s	450	10	3.7.3

4 simulations from chapters 2 and 3 are taken as examples. The simulation lengths and number of run are described in Table B.1.

Figure B.1 shows and example of convergence of the mean cell throughput within one run. The example is characteristic of the convergence of the mean cell throughput within a run of 10s. Generally, in the last 2s of the run, the cumulated mean cell throughput oscillate within a "box" of only 0.02Mbps amplitude.

Figure B.2 shows the empirical CDFs of the different per run mean cell throughputs (instantiated from the different X_{PS}^is) for each of the four simulations described in Table B.1. It is noticeable here that the different per run mean cell throughputs span over intervals of at least 4 Mbps, it is therefore important to generate a large number of runs to obtain a high precision. The precision of the average cell throughput estimation is assessed in figure B.3. It shows that the s values for the different simulations are very small (< 0.2Mbps) compared to the estimated average cell throughput (which is always in the order of 10Mbps).

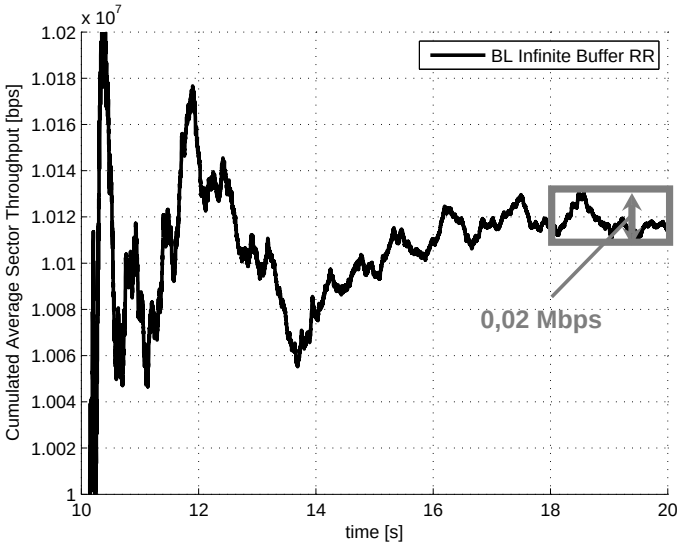

Figure B.1: cumulated mean throughput within one run for the RR PS algorithm simulated with the infinite buffer. the cumulated mean throughput at time t is defined by the mean of all the instant cell throughputs values until time t.

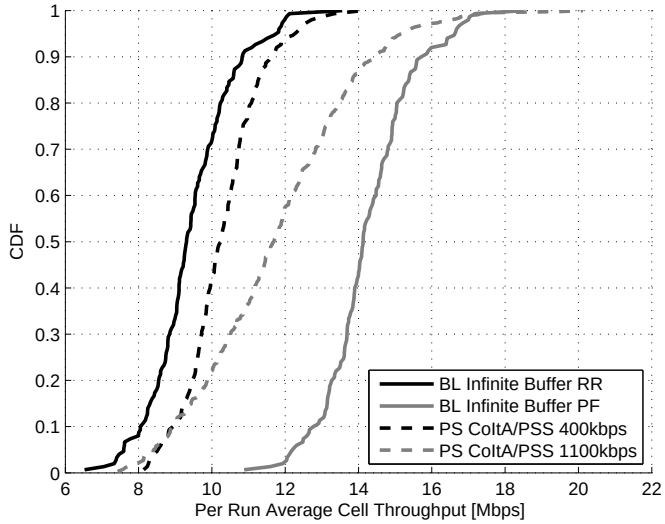

Figure B.2: Empirical CDFs of the different per run mean cell throughputs.

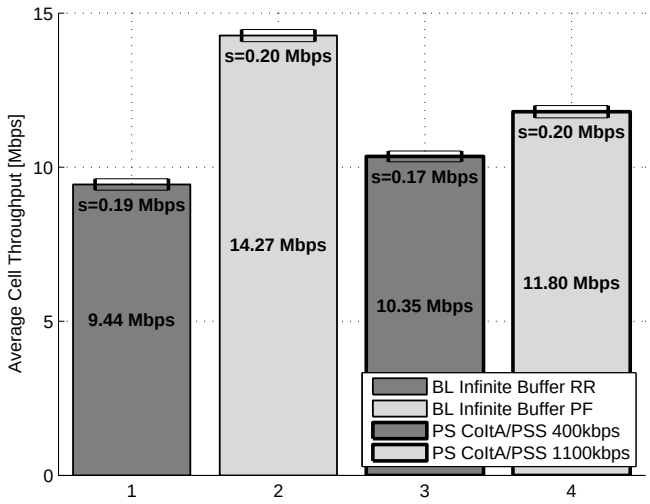

Figure B.3: Estimated average cell throughput with the precision intervals derived from (B.4)

B.2.2 Coverage

For the coverage, it is necessary to reason on UE throughput basis. The UE throughput, similarly to the mean cell throughput in (B.2), can be expressed as a random variable:

$$\begin{aligned} \mathbf{Y}_{PS}: \quad & \Omega \rightarrow \mathbb{R}^{N_{UE}} \\ & \omega \rightarrow \mathbf{y} \end{aligned} \tag{B.5}$$

every instantiation of Ω corresponds to N_{UE} different UEs and therefore to N_{UE} different throughput values in vector $\mathbf{y}$. $\mathbf{Y}_{PS}$ can be relaxed into N_{UE} different random variables:

$$\begin{aligned} \mathbf{Y}_{PS}[i]: \quad & \Omega \rightarrow \mathbb{R}^{N_{UE}} \\ & \omega \rightarrow \mathbf{y}[i] \end{aligned} \tag{B.6}$$

The different $\mathbf{Y}_{PS}[i]$ are identically distributed as they are inter changeable among each other. However, they are not necessarily independent from each other. Indeed, for example, $\mathbf{y}[1]$ and $\mathbf{y}[2]$ are from UE 1 and 2 and their throughput are generated depending on the same set of other UEs present in the system.

F is defined as the CDF of $\mathbf{Y}_{PS}[i]$ (where i does not matter as the different $\mathbf{Y}_{PS}[i]$ are identically distributed) and the coverage is defined as:

$$Cov = F^{-1}(0.05) \tag{B.7}$$

In the thesis, the throughput CDF is estimated by a scale function where the different steps are equal to:

$$\hat{F}(\mathbf{Ry}[k]) = \frac{k}{M} \tag{B.8}$$

where:

- $\mathbf{Ry}$ is the vector containing all the throughput samples generated during the simulation throughout all runs and by each UE. Samples are ranked in the vector by increasing values.

- M is the total number of generated UE throughput samples

As $\hat{F}$ considers all samples from each run, it is a good estimation only if the different $\mathbf{Y}_{PS}[i]$ are independent. It is simply assumed here that they are independent enough.

Following (B.8) the coverage is estimated as follows:

$$\hat{Cov} = \hat{F}^{-1}(0.05) = \hat{F}^{-1}\left(\frac{k_{0.05}}{M}\right) \tag{B.9}$$

In order to assess the validity of the estimate, the distribution of $\hat{F}(x)$ is observed:

$$P\left(\hat{F}(x) = \frac{k}{M}\right) = C_M^k \cdot F(x)^k \cdot (1 - F(x))^{M-k} \tag{B.10}$$

$M \cdot \hat{F}(x)$ is binomially distributed (and with M large enough can be approximated by a normal distribution) and therefore:

$$E\left(\hat{F}(x)\right) = F(x) \tag{B.11}$$

$$Std\left(\hat{F}(x)\right) = \frac{\sqrt{F(x) \cdot (1 - F(x))}}{\sqrt{M}} \tag{B.12}$$

This means that all the points of the empirical throughput CDF can correspond to the actual coverage with a certain probability. It is chosen here to define a precision interval for the coverage with all the points of the empirical throughput CDF that have at least a certain likelihood of including the actual coverage $Cov = F^{-1}(0.05)$. This set of points is defined by the points x for which the empirical CDF image is no further than two standard deviations from 0.05:

$$\left|\hat{F}(x) - 0.05\right| < 2 \cdot Std\left(\hat{F}(Cov)\right) \tag{B.13}$$

In the infinite buffer simulations run in the thesis, there is always $M = 4500$. In that case: $2 \cdot Std\left(\hat{F}(Cov)\right) \approx 0.007$. Therefore, the precision interval is $\left[\hat{F}^{-1}(0.043)\, \hat{F}^{-1}(0.057)\right]$. Figure B.4 show the coverage with the precision interval of the simulations of Table B.1. The value s corresponds to the half-size of the precision interval. It is shown that the simulations are very precise in term of coverage as the precision is of the order of 1%, except for the RR simulation where it approaches 10%.

B.3 finite buffer

For finite buffer simulations, the statistical relevance has been assessed empirically with visual methods. it is indeed not possible to create a clear theoretical framework as for infinite buffer simulations. As UEs come in and out of the network during a simulation, each UE is successively influenced by the different UE changes in the network. Only one "run" is used for these simulations as the UEs and more generally the network conditions are changing constantly. It is assumed that at some point, the cumulated mean cell throughput converges. Therefore, the main focus has been to visually look at the convergence of the cumulated mean cell throughput. Generally it has been verified in every simulation that the cumulated mean cell throughput oscillate with an amplitude smaller than 0.1Mbps. This is further shown Figures B.5, B.6, B.7, B.8 and B.9 that show time

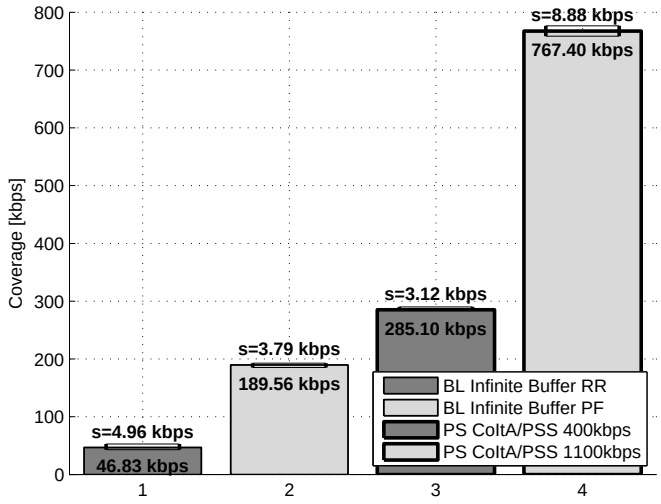

Figure B.4: Coverage with precision interval for infinite buffer simulations.

Table B.2: Finite buffer simulation for evaluation of statistical significance

Simulation	Sim length	N_{UE}	# Sim. Sectors	# UE finished	Section
BL RR	500s	30	3	3056	2.6.5
BL PF	500s	30	3	5065	2.6.5
FL BFF = 0.5 BQPP	40s	10	57	9331	4.2.6

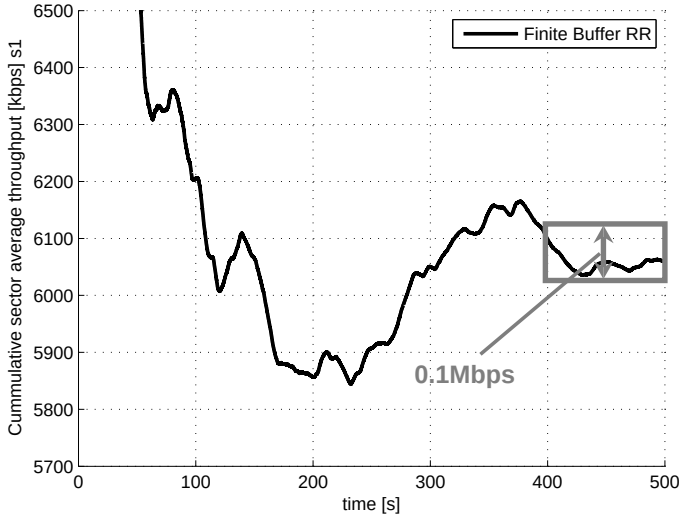

Figure B.5: time versus the cumulated mean cell throughput for sector 1 of the base line finite buffer simulation with RR

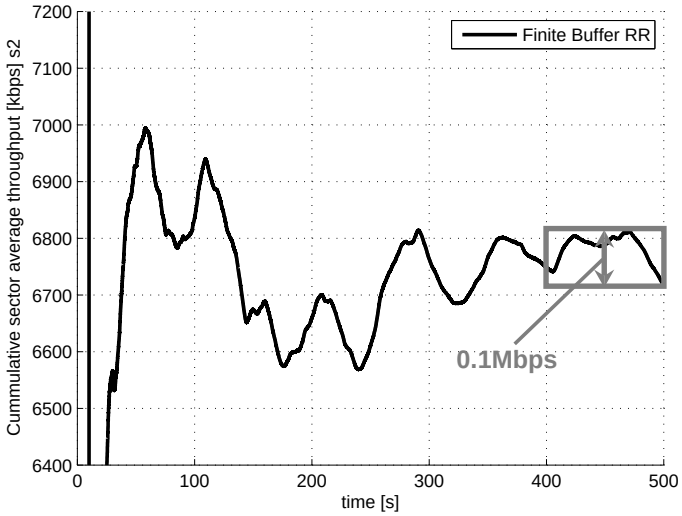

Figure B.6: time versus the cumulated mean cell throughput for sector 2 of the base line finite buffer simulation with RR

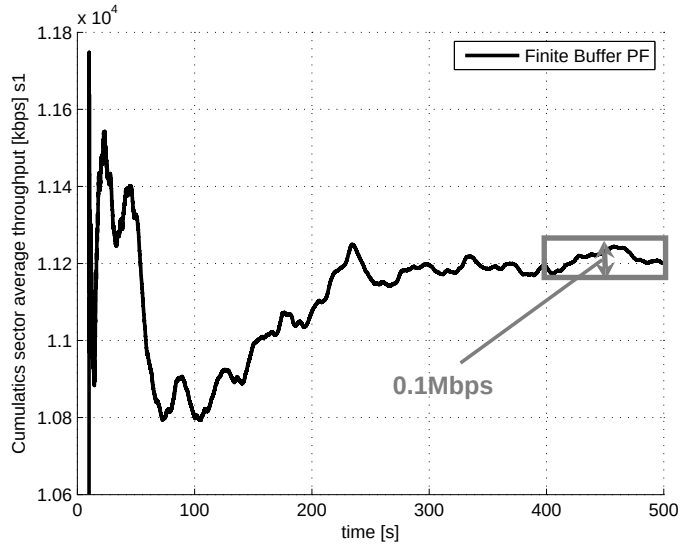

Figure B.7: time versus the cumulated mean cell throughput for sector 1 of the base line finite buffer simulation with PF

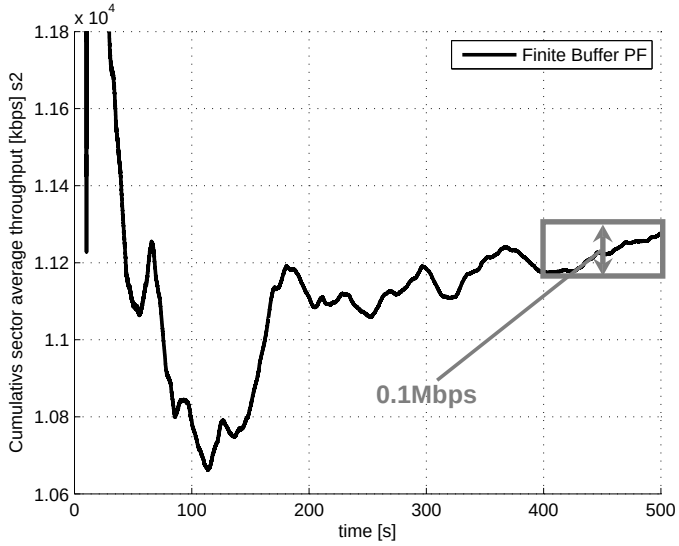

Figure B.8: time versus the cumulated mean cell throughput for sector 2 of the base line finite buffer simulation with PF

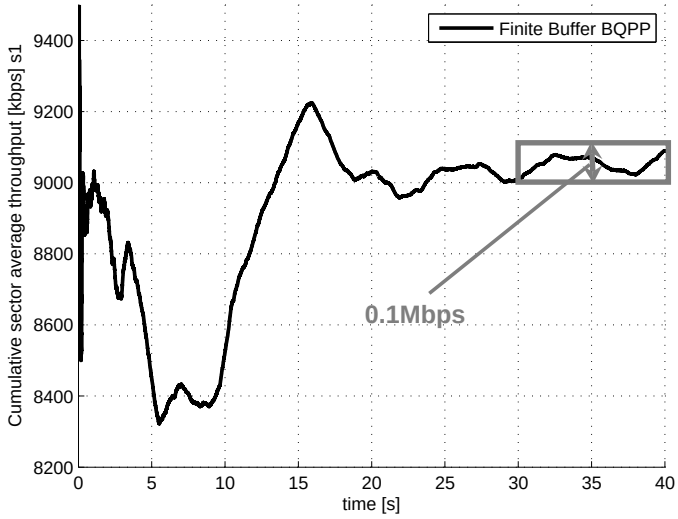

Figure B.9: time versus the cumulated mean cell throughput for sector 1 of the fractional load finite buffer simulation with BFF=0.5 and BQPP

versus the cumulated mean cell throughput for some simulated sector for the set of simulations described in Table B.2.

The average cell throughput is the sample mean of the different final cumulated mean cell throughputs in different simulated sectors. Figure B.10 shows the empirical CDF of the final cumulated mean cell throughputs for the simulations described in Table B.2. The important point is that in all simulations, the lowest and highest mean throughput are separated by around 0.5Mbps which is higher than the oscillation amplitude of the cumulated mean throughput per sector at the end of the simulation shown previously. The important message is that the cell throughput should not be relied on with a precision higher than 0.5Mbps (around 5%).

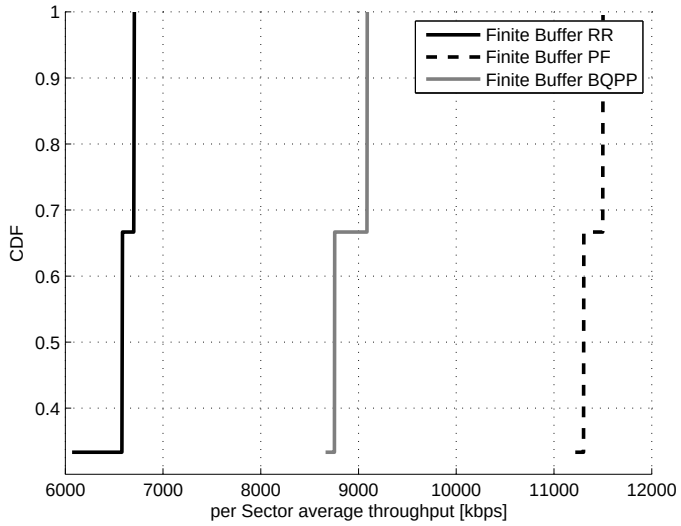

Figure B.10: empirical CDF of the final cumulated mean cell throughputs for finite buffer simulations.

As for the coverage, it is not possible either to create a clear theoretical framework. Therefore, the great number of sessions is relied on to provide a precise estimate.

Bibliography

[1] H. Yang, "A Road to Future Broadband Wireless Access: MIMO-OFDM-Based Air Interface," *IEEE Communications Magazine*, vol. 43, no. 1, pp. 53–60, January 2005.

[2] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian, and V. Niemi, Eds., *UMTS Networks, Architechture, Mobility and Services*, 2nd ed. John Wiley & Sons, 2005.

[3] H. Holma and A. Toskala, Eds., *WCDMA for UMTS, HSPA Evolution and LTE*, 4th ed. John Wiley & Sons Ltd, 2007.

[4] T. Halonen, J. Romero, and J. Melero, Eds., *GSM, GPRS and EDGE Performance, Evolution towards 3G/UMTS*. John Wiley & Sons, 2002.

[5] F. Adachi, "Evolution Towards Broadband Wireless Systems," in *Proceedings of the International Symposium on Wireless Personal Multimedia Communications (WPMC)*, vol. 1, Honolulu, Hawaii, October 2002, pp. 19–26.

[6] H. Yin and S. Alamouti, "OFDMA: A Broadband Wireless Access Technology," *IEEE Sarnoff Symposium*, March 2006.

[7] *Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation*, WiMAX Forum, www.wimaxforum.org/home, August 2006.

[8] 3GPP Technical Report 25.913, version 7.3.0, *Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)*, March 2006.

[9] 3GPP Technical Report 25.814, version 7.1.0, *Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA) (Release 7)*, September 2007.

[10] A. Pokhariyal, T. E. Kolding, and P. E. Mogensen, "Performance of Downlink Frequency Domain Packet Scheduling for the UTRAN Long Term Evolution," in *IEEEProceedings of the Personal, Indoor and Mobile Radio Communications*, Helsinki, Finland, September 2006.

[11] 3GPP Technical Specification 23.401, version 8.2.0, *General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 8)*, June 2008.

[12] 3GPP Technical Specification 23.402, version 8.2.0, *Architecture Enhancements for non 3GPP Accesse (Release 8)*, June 2008.

[13] 3GPP Technical Specification 36.300, version 8.5.0, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 8)*, May 2008.

[14] D. Soldani, M. Li, and R. Cuny, Eds., *QoS and QoE Management in UMTS Cellular Systems*. John Wiley & Sons Ltd, 2006.

[15] 3GPP Technical Specification 23.107, version 7.1.0, *Quality of Service (QoS) concept and architecture (Release 7)*, September 2007.

[16] 3GPP Technical Specification 23.203, version 8.2.0, *Policy and Chargin Control Architecture (Release 8)*, June 2008.

[17] R. Chang, "Synthesis of Band-Limited Orthogonal Signals for Multichannel Data Transmission," *Bell Systems Technical Journal*, vol. 45, pp. 1775–1796, December 1966.

[18] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, April 1980, pp. 964–967.

[19] S. Weinstein and P. Ebert, "Data Transmission by Frequency-Division Multiplexing Using the Discrete Fourier Transform," *IEEE Transactions on Communications*, vol. 1, pp. 628–634, October 1971.

[20] J. A. C. Bingham, "Multicarrier Modulation For Data Transmission: An Idea Whose Time Has Come," *IEEE Communications Magazine*, vol. 28, pp. 5–14, May 1990.

[21] 3GPP Technical Specification 36.211, version 8.3.0, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Physical channels and modulation (Release 8)*, May 2008.

[22] 3GPP Technical Specification 36.104, version 8.2.0, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Base Station (BS) radio transmission and reception (Release 8)*, May 2008.

[23] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," *IEEE Proceedings of the International Conference on Communications (ICC)*, vol. 1, pp. 331–335, June 1995.

[24] M. Alasti, F. R. Farrokhi, M. Olfat, and K. J. R. Liu, "Service level agreement (SLA) based scheduling algorithms for wireless networks," *IEEE Proceedings of the International Conference on Communications (ICC)*, vol. 2, pp. 1028–1032, June 2004.

[25] F. R. Farrokhi, M. Olfat, M. Alasti, and K. J. R. Liu, "Scheduling Algorithms for Quality of Service Aware OFDMA Wireless Systems," in *IEEE Proceedings of the Global Telecommunications Conference (GLOBECOM)*, Dallas, Texas, USA, November 2004, pp. 2689–2693.

[26] A. Parekh, "A Generalized Processor Sharing Approach to Flow Control in Integrated Service Networks," Ph.D. dissertation, MIT - Laboratory for Information and Decision Systems, Cambridge, Massachusetts, 1992.

[27] L. Songwu, V. Bharghavan, and R. Srikant, "A Wireless Fair Service Algorithm for Packet Cellular Networks," *ACM Proceedings of the International Conference on Mobile Computing and Networking (MOBICOM)*, October 1998.

[28] ——, "Fair scheduling in wireless packet networks," *IEEE Transactions on Networking*, vol. 7, pp. 473–489, August 1999.

[29] J. H. Rhee, T. H. Kim, and D. K. Kim, "A Wireless Fair Scheduling Algorithm for 1xEV-DO System," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 2, Atlantic City, New Jersey, USA, October 2001, pp. 743–746.

[30] K. Khawam and D. Kofman, "Opportunistic Weighted Fair Queueing," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, September 2006.

[31] C. Lengoumbi, P. Martins, and P. Godlewski, "An Opportunist Extension of Wireless Fair Service for Packet Scheduling in OFDMA," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, April 2007.

[32] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, January 1997.

[33] P. A. Hosein, "QoS Control for WCDMA High Speed Packet Data," in *IEEEProceedings of the International Workshop on Mobile and Wireless Communication Networks (MWCN)*, Stockholm, Sweden, September 2002, pp. 169–173.

[34] A. Jalali, R. Padovani, and R. Pankaj, "Data Throughput of CDMA-HDR High Efficiency-High Data Rate Personal Communication Wireless System," *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 3, pp. 1854–1858, May 2000.

[35] J. M. Holtzman, "CDMA Forward Link Water Filling Power Control," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 3, Tokyo, Japan, May 2000, pp. 1663–1667.

[36] ——, "Asymptotic Analysis of Proportional Fair Algorithm," in *IEEEProceedings of the Personal, Indoor and Mobile Radio Communications*, vol. 2, San Diego, California, USA, September 2001, pp. F33–F37.

[37] M. Lundevall, B. Olin, J. Olsson, N. Wiberg, S. Wänstedt, J. Eriksson, and F. Eng, "Streaming Applications over HSDPA in Mixed Service Scenarios," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 2, Los Angeles, California, USA, September 2004, pp. 841–845.

[38] T. E. Kolding, "QoS-Aware Proportional Fair Packet Scheduling with Required Activity Detection," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, Montreal, Canada, 2006.

[39] P. A. Hosein, "Scheduling of VoIP traffic over a time-shared wireless packet data channel," in *IEEEProceedings of the International Conference on Personal Wireless Communications (PWC)*, January 2005, pp. 38–41.

[40] A. L. Stolyar and K. Ramanan, "Largest Weighted Delay First Scheduling: Large Deviations and Optimality," *The (IMS) Annals of Applied Probabilities*, vol. 11, pp. 1–48, 2001.

[41] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "CDMA Data QoS Scheduling on the Forward Link with Variable Channel Conditions," *Bell Systems Technical Memo*, 2000.

[42] P. Ameigeiras, J. Wigard, and P. Mogensen, "Performance of the M-LWDF scheduling algorithm for streaming services in HSDPA," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 2, Los Angeles, California, USA, September 2004, pp. 999–1003.

[43] A. Braga, E. Rodrigues, and F. Cavalcanti, "Packet Scheduling for VOIP Over HSDPA in Mixed Traffic Scenarios," in *IEEEProceedings of the Personal, Indoor and Mobile Radio Communications*, September 2006.

[44] G. S. Y. Li;, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 127–143, December 2005.

[45] T. Nguyen and Y. Han, "A Proportional Fairness Algorithm with QoS Provision in Downlink OFDMA Systems," *IEEE Communications Letters*, vol. 10, pp. 760–762, November 2006.

[46] M. Kaneko, "Medium Access Techniques for Multi-Carrier Wireless Systems," Ph.D. dissertation, Aalborg University – Department of Communication Technology, Aalborg, Denmark, November 2007.

[47] C. Wengerter, J. Ohlhorst, and A. G. E. v. Elbwart, "Fairness and Throughput Analysis for Generalized Proportional Fair Frequency Scheduling in OFDMA," *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 3, pp. 1903–1907, May 2005.

[48] J. Liebeherr, D. Wrege, and D. Ferrari, "Exact admission control for networks with a bounded delay service," *IEEE Transactions on Networking*, vol. 4, pp. 885–901, December 1996.

[49] P. A. Hosein, "A Class-Based Admission Control Algorithm for Shared Wireless Channels Supporting QoS Services," in *Proceedings of the Fifth IFIP TC6 International Conference on Mobile and Wireless Communications Networks*, October 2003.

[50] K. Pedersen, "Quality based HSDPA access algorithms," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 4, Los Angeles, California, USA, September 2004, pp. 2498–2502.

[51] S. S. Jeong, J. A. Han, and W. S. Jeon, "Adaptive connection admission control scheme for high data rate mobile networks," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 4, Dallas, Texas, USA, September 2005, pp. 2607–2611.

[52] D. Niyato and E. Hossain, "Connection Admission Control Algorithms for OFDM Wireless Networks," in *IEEE Proc. GLOBECOM Conf.*, December 2005.

[53] C. F. Ball, E. Humburg, K. Ivanov, and F. Treml, "Performance Analysis of IEEE802.16 Based Cellular MAN with OFDM-256 in Mobile Scenarios," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, May 2005.

[54] M. Maqbool, M. Coupechoux, and P. Godlewski, "Comparison of Various Frequency Reuse Patterns for WiMAX Networks with Adaptive Beamforming," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, Singapore, May 2008, pp. 2582–2586.

[55] Huawei, "Further analysis of soft frequency reuse scheme," *3GPP TSG-RAN1, R1-050841*, September 2005.

[56] Ericsson, "Inter-cell interference handling for E-UTRA," *3GPP TSG-RAN1, R1-050764*, September 2005.

[57] M. Pischella and J.-C. Belfiore, "Achieving a Frequency Reuse Factor of 1 in OFDMA cellular networks with cooperative communications," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, Singapore, May 2008, pp. 653–657.

[58] 3GPP Technical Specification 45.005, version 5.4.0, *Radio transmission and reception*, June 2002.

[59] W. C. Jakes, Ed., *Microwave Mobile Communications*. ISBN 0-7803-1069-1: IEEE Press, New Jersey, 1974.

[60] 3GPP Technical Specification 36.214, version 8.3.0, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Physical layer - measurements (Release 8)*, May 2008.

[61] 3GPP Technical Specification 36.321, version 8.2.0, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Medium Access Control (MAC) protocol specification (Release 8)*, May 2008.

[62] 3GPP Technical Specification 36.322, version 8.2.0, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Radio Link Control (RLC) protocol specification (Release 8)*, May 2008.

[63] K. Brueninghaus, D. Astley, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link Performance Models for System Level Simulations of Broadband Radio Access Systems," in *IEEEProceedings of the Personal, Indoor and Mobile Radio Communications*, vol. 4, Berlin, Germany, September 2005, pp. 2306–2311.

[64] D. Lopez, C. Ubeda, I. Kovacs, F. Frederiksen, and K. Pedersen, "Performance of Downlink UTRAN LTE under Control Channel Constraints," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, Singapore, May 2008, pp. 2512–2516.

[65] F. Frederiksen and T. E. Kolding, "Performance and Modeling of WCDMA/HSDPA Transmission/H-ARQ Schemes," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, vol. 1, Vancouver, Cananda, September 2002, pp. 472–476.

[66] K. I. Pedersen, G. Monghal, I. Z. Kovacs, T. E. Kolding, A. Pokhariyal, F. Frederiksen, and P. E. Mogensen, "Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback," in *Proceedings of IEEE Vehicular Technology Confreence (VTC)*, Baltimore, USA, October 2007.

[67] M. Nakamura, Y. Awad, and S. Vadgama, "Adaptive Control of Link Adaptation for High Speed Downlink Packet Access (HSDPA) in W-CDMA," in *IEEEProceedings of the Wireless Personal Multimedia Communications Conference (WPMC)*, vol. 2, Honolulu, Hawaii, October 2002, pp. 382–386.

[68] A. Pokhariyal, K. I. Pedersen, G. Monghal, I. Z. Kovacs, C. Rosa, T. E. Kolding, and P. E. Mogensen, "HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness for the UTRAN Long Term Evolution," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, Dublin, Ireland, April 2007, pp. 2761–2765.

[69] Nortel, "DL Performance Evaluation for E-UTRA," *3GPP TSG-RAN1, R1-071952*, May 2007.

[70] Motorola, "LTE Downlink System Performance Verification Results," *3GPP TSG-RAN1, R1-071976*, April 2007.

[71] B. Classon, K. Baum, V. Nangia, R. Love, Y. Sun, R. Nory, K. Stewart, A. Ghosh, and R. Ratasuk, "Overview of UMTS Air-Interface Evolution," in *IEEE Proceedings of the Vehicular Technologie Conference (VTC)*, Montréal, Québec, Canada, September 2006.

[72] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-part I: theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, March 2005.

[73] ——, "Cross-layer optimization for OFDM wireless networks-part II: algorithm development," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625–634, March 2005.

[74] A. Pokhariyal, G. Monghal, K. I. Pedersen, P. E. Mogensen, I. Z. Kovacs, C. Rosa, and T. E. Kolding, "Frequency Domain Packet Scheduling Under Fractional Load for the UTRAN LTE Downlink," in *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, Dublin, Ireland, April 2007, pp. 699–703.

[75] P. E. Mogensen, W. Na, I. Z. Kovacs, F. Frederiksen, A. Pokhariyal, K. I. Pedersen, K. Hugl, and M. Kuusela, "LTE Capacity versus Shannon," in *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, Dublin, Ireland, April 2007, pp. 1234–1238.