ADMISSION CONTROL AND RESOURCE ALLOCATION FOR LTE UPLINK SYSTEMS

Oscar Delgado

A thesis

in

The Department

of

Electrical Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Electrical Engineering
Concordia University
Montréal, Québec, Canada

August 2010

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-70964-1
Our file Notre référence
ISBN: 978-0-494-70964-1

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By : **Oscar Delgado**

Entitled : **Admission Control and Resource Allocation for LTE Uplink systems**

and submitted in partial fulfilment of the requirements for the degree of

Master of Electrical Engineering

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee :

______________________________ Chair
Dr. Dongyu Qiu

______________________________ External Examiner
Dr. Clement Lam

______________________________ Examiner
Dr. Yousef Shayan

______________________________ Supervisor
Dr. Brigitte Jaumard

Approved by ______________________________
Dr. Mojtaba Kahrizi
Graduate Program Director

____
Dr. Robin Drew
Dean of the Faculty of Engineering and Computer Science

Date ______________________________

Abstract

Admission Control and Resource Allocation for LTE Uplink systems

Oscar Delgado

Long Term Evolution (LTE) radio technologies aim not only to increase the capacity of mobile telephone networks, but also to provide high throughput, low latency, an improved end-to-end Quality of Service (QoS) and a simple architecture. The Third Generation Partnership Project (3GPP) has defined Single Carrier FDMA (SC-FDMA) as the access technique for the uplink and Orthogonal Frequency Division Multiple Access (OFDMA) for the downlink.

It is well known that scheduling and admission control play an important role for QoS provisioning, and that they are strongly related. Knowing that we can take full advantage of this property we can design an admission control mechanism that uses the design criterion of the scheduling scheme.

In this thesis, we developed two new algorithms for handling single-class resource allocation and two algorithms for handling multi-class resource allocation, as well as a new admission control scheme for handling multi-class Grade of Service (GoS) and QoS in uplink LTE systems. We also present a combined solution that uses the resource allocation and the admission control properties to satisfy the GoS and QoS requirements.

System performance is evaluated using simulations. Numerical results show that the proposed scheduling algorithms can handle multi-class QoS in LTE uplink systems with a little increase in complexity, and can be used in conjunction with admission control to meet the LTE requirements. In addition, the proposed admission control algorithm gain for the most sensitive traffic can be increased without sacrificing the overall system capacity. At the same time, guaranteeing GoS and maintaining the basic QoS requirements for all the admitted requests.

Acknowledgments

This thesis proved to be a challenging task as I started to learn about wireless technologies after getting impressed from its rising popularity in the world. Deciding about the specific topic of my research work was not an easy task. The contributions of many people, in different ways, have made this possible. I would like to extend my gratitude to the following.

Thank God for the wisdom and perseverance that he has been bestowed upon me during this time, and indeed, throughout my life.

I would like to express my sincere gratitude to my supervisor Dr. Brigitte Jaumard who provided me guidelines on how to select my research topic. I really appreciate her precious recommendations which helped me a lot in moving ahead with my research.

I would also like to express my gratitude to my friends for their continue support and encouragement. More than anyone else, I would like to thank my parents, Mr. Humberto and Mrs. Maria and my two sisters Yenny and Cathia. Without their confidence in my abilities, I would not have made it through this program.

Contents

List of Figures

List of Tables

List of Abbreviations

3GPP	Third Generation Partnership Project
4G	Fourth Generation
AC	Admission Control
CDMA	Code Division Multiple Access
EPC	Evolved Packet Core
EPS	Evolved Packet System
E-UTRAN	Evolved-UTRAN
FDD	Frequency Division Duplexing
FFT	Fast Fourier Transform
FDMA	Frequency Division Multiple Access
FTP	File Transfer Protocol
GBR	Guaranteed Bit Rate
GoS	Grade of Service
HSS	Home Subscription Server
IETF	Internet Engineering Task Force
IP	Internet Protocol
LTE	Long Term Evolution
MAC	Medium Access Control
MME	Mobility Management Entity
OFDMA	Orthogonal Frequency Division Multiple Access

PAPR	Peak to Average Power Ratio
PCRF	Policy and Charging Resource Function
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PDU	Payload Data Unit
P-GW	Packet Data Network Gateway
QCI	QoS Class Identifier
QoS	Quality of Service
RB	Resource Block
RLC	Radio Link Control
RRC	Radio Resource Control
SAE	System Architecture Evolution
SAP	Service Access Point
SC-FDMA	Single Carrier Frequency Division Multiple Access
S-GW	Serving Gateway
SMTP	Simple Mail Transfer Protocol
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UE	User Equipment
UTRAN	UMTS Terrestrial Radio Access Network
VoIP	Voice over IP
WiMAX	Worldwide Interoperability for Microwave Access

Chapter 1

Introduction

1.1 Generalities

Wireless technologies aim to provide broadband access and services anytime, anywhere using any terminal. The possibilities of voice, video, and data services over wireless transmission seem endless, but offering such services poses many challenges involving efficient use of resources, quality of service, etc.

Streaming media is becoming a pivotal role in data services. In addition, a greater range of services like: online gaming, real-time video, voice services, streaming media, Web browsing, and non-real-time services, i.e., FTP, SMTP, have different requirements of network performance such as bandwidth, delay, jitter, and security. Streaming media closely bonds the requirements of subscribers and the provisioning capability of networks. As a key broadband data service, it will consume the major share of bandwidth resources.

The wireless market growth is stimulated by two main factors: subscribers upgrading their mobile devices to have the newest, most advanced features, and subscriber growth in emerging countries. It is expected that by 2013, there will be more than 1 billion mobile devices that can connect to the Internet. That includes phones, gaming consoles, netbooks, eBook readers, GPS systems and car navigation systems. The majority of these customers will be served by Fourth Generation (4G)

Technologies.

Long Term Evolution (LTE) aims to meet the needs of future broadband cellular communications. The standard is based on a decentralized architecture, with scheduling and admission control functionality embedded in evolved Node-B at layer 3, so it can utilize local cell measurement information to take the admission control decision. LTE also provides Quality of Service (QoS) to the different types of services.

1.2 Problem Statement

Radio Resource Management role is to ensure that the radio resources are efficiently used, optimizing in this way the system capacity and the end user performance. The most important algorithms in LTE are admission control, scheduling, power control and interference control. These network algorithms are not specified by the Third Generation Partnership Project (3GPP) so it gives the opportunity to network vendors and operators to design and tune the algorithms according to their needs.

The eNodeB admission control algorithm decides whether new requests in the cell are granted or rejected. Admission control takes into account the resource situation in the cell, the QoS requirements of the new request, as well as the priority levels, and the currently provided QoS to the active sessions in the cell. Thus, an admission control algorithm aims at only admitting new request up to the point where the scheduler in the cell can converge to a feasible solution where the promised QoS requirements are fulfilled for at least all the requests with high priority.

It is clear that uplink has become very important due to the increasing throughput demands upon the uplink channels. This thesis will first analyze and develop new scheduling algorithms for single-class and multi-class environments focusing in meeting the QoS requirements for throughput and delay. Then, we will focus on joint admission control and scheduling schemes. We will propose a novel algorithm for handling the priorities while fulfilling the QoS objective of all granted requests. To design such an algorithm, we will present a combined solution that uses the resource allocation and the admission control properties to satisfy our objective.

1.3 Scope of Thesis Work

This thesis main concern is on uplink admission control and scheduling solutions. It will not focus on downlink although all the properties of uplink can hold for downlink, but this is not necessarily true if we apply the downlink solutions into the uplink, mainly because of the additional constraints in the uplink.

This research thesis uses simulation tools to clarify the trade-offs between some key measures in admission control, such as system throughput, blocking and dropping probabilities. All these measures are considered in a single-cell SC-FDMA environment with convenient admission control and scheduling algorithms designed for this purpose.

1.4 Contributions of the Thesis

Our approach, dealing with scheduling and admission control for LTE uplink systems, is to propose scheduling algorithms that supports QoS, where the main focus is on throughput and delay, for single-class and multi-class systems. We also propose an admission control scheme that works in conjunction with our new scheduling algorithms, to guarantee quality of service and grade of service. To this end, we have made the following contributions.

- In the first part of our research [DJ10c], we developed an optimization model that, for the first time, includes the LTE uplink requirements, i.e., block contiguity constraint. We also propose two algorithms that assign resources in a way that maximizes the total sum throughput while meeting the delay requirement and at the same time trying to be fair among all the users. Remember that the throughput is dependent on the distance and thus users close to the base station tend to have higher throughput. We demonstrate through simulations that our proposed algorithms outperform the most recently proposed algorithms by introducing the delay as part of the maximization problem.

- In [DJ10b], we extended our optimization model to support multiple classes of service, as, in practice, LTE defines until nine different types of services. Here, we focus on the proportional fair approach to develop two algorithms that support multi-class, meet the QoS requirements in terms of throughput and delay, and have a fair behavior among the users. In the simulation part, we found that with both algorithms, we can have good performance even though we are introducing new requirements.

- Finally in [DJ10a], we developed a novel admission control algorithm that in conjunction with our scheduling algorithm for multi-class environments, supports QoS and Grade of Service (GoS) for new and existing requests. This solution adaptively adjusts the throughput and delay according to the traffic load, assigning resources in a fair way. We evaluated the performance of our proposed solution in comparison with the basic admission control criteria. Numerical results show that it is possible to increase the system capacity and at the same time fulfill the QoS and GoS targets.

1.5 Thesis Outline

Technical background of Long Term Evolution is discussed in Chapter 2, covering different technical aspects of LTE like downlink and uplink multiple access schemes, and technical specifications on layers 1-3.

Chapter 3 makes a comprehensive review of the previous work found in the literature for scheduling/resource allocation and admission control schemes for LTE systems and for other wireless technologies.

Chapter 4 develops the mathematical model for LTE uplink scheduling, which includes the uplink block contiguity constraints and the delay requirement. It also covers the algorithm design of scheduling for single-class and multi-class systems, and then a joint admission control and scheduling solution for multi-class systems.

Chapter 5 contains the simulation results, followed by a critical analysis. Conclusion and future work are discussed in the last chapter.

Chapter 2

LTE Technical Background

This section contains a short introduction of the most important technical aspects of Long Term Evolution. More details about LTE can be found in [3GP10a], [DPSB08], [HA09], and [STB09].

2.1 Overview

As wireless technologies evolved, the access techniques used also exhibited increase in efficiency. The first generation of wireless technologies used time division multiple access (TDMA) and frequency division multiple access (FDMA). In second generation systems, one set of standards used the combination of FDMA and TDMA and the other set introduced code division multiple access (CDMA). CDMA enables the third generation systems. One critical issue with CDMA is that it suffers from low spectral flexibility and computationally intensive time-domain equalization for wideband channels.

Recently, new access schemes like Orthogonal FDMA (OFDMA) and Single Carrier FDMA (SC-FDMA) are introduced for the Four generation systems. They are based on efficient Fast Fourier Transform (FFT) algorithms and frequency domain equalization. They also make possible to control the bandwidth and form the spectrum in a flexible way. However, they require advanced channel allocation.

In this context, LTE aims to provide increased data rates, improved spectral efficiency, and reduced user and control plane latency. To accomplish these main goals, LTE has chosen Orthogonal Frequency Division Multiple Access (OFDMA) for the downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) for the uplink.

These multiple access techniques provide orthogonality between the sub-channels, reducing in this way the interference and improving the total network capacity. OFDMA was chosen for the downlink due to its high data rate capacity and its high spectral efficiency. SC-FDMA was chosen for the uplink for its lower Peak to Average Power Ratio (PAPR) to maximize battery life in mobile devices.

LTE technology has been designed to support packet switched services, in contrast with previous mobile systems which are circuit-switched based. It aims to provide full Internet Protocol (IP) connectivity between the user's mobile device and the Packet Data Network (PDN), without any disruption to the end user's applications during mobility.

The main performance targets for LTE can be summarized as follows:

- Increased user data rates.
- Increased cell-edge bit-rate, for uniformity of service provision.
- Reduced delays, in connection establishment and transmission latency.
- Reduced cost per bit.
- Flexibility of spectrum usage.
- Simplified network architecture.
- High level of mobility and security.
- Optimized power consumption for the mobile terminal.

To address these objectives, LTE needs not only the evolution of the radio access techniques through the Evolved-UTRAN (E-UTRAN) but also the evolution of the non-radio aspects, named System Architecture Evolution (SAE), which includes the Evolved Packet Core (EPC). LTE and SAE comprise the Evolved Packet System (EPS).

2.2 LTE Network Architecture

In this section, we present the overall LTE network architecture, giving an overview of the main functions provided by the core network and the access network. Figure 1 shows the architecture and network elements. The logical nodes and connections shown therein represent the basic system architecture configuration.

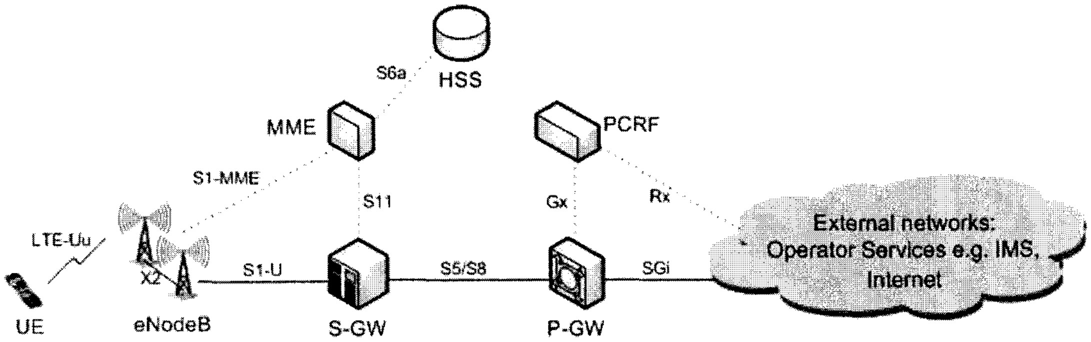

Figure 1: LTE Network architecture

The eNodeB is in charge of single cell radio resource management decisions, scheduling of users in downlink and uplink, handover decisions, etc. The eNodeB is connected to the core network using the S1 interface.

The Mobility Management Entity (MME) is the control node which processes the signaling between the User Equipment (UE) and the core network. The main functions supported by the MME are: Authentication and security, which ensures that the UE is who it claims to be, and Mobility Management, which takes care of keeping track of the location of all UE's in its service area, managing subscription profile and service connectivity. The MME stores this information for the duration it is serving the UE.

The serving gateway (S-GW) serves as the local mobility anchor for the data bearers when the UE moves among eNodeB's. The high level function of S-GW is UP tunnel management and switching.

Home Subscription Server (HSS) stores user's subscription information such as the QoS profile and any access restriction, i.e., roaming. It also holds information about the networks to which the user can connect.

The Packet Data Network Gateway (P-GW) is the edge router between the LTE network and the external packet data networks. It is responsible for IP address allocation for the UE, as well as QoS enforcement and flow-based charging according to rules from the Policy and Charging Resource Function (PCRF). It also serves as the mobility anchor for inter-working with non-3GPP technologies.

PCRF is the network element that is responsible for policy control decision-making and Charging Control. It also makes decisions on how to handle the services in terms of QoS.

2.3 Physical Layer

LTE physical layer is based on the use of OFDMA and SC-FDMA technologies and is characterized by the design principle of resource usage based on dynamically allocated shared resources rather than having dedicated resources reserved for a single user [3GP10b].

Physical layer provides data transport services to higher layers with the help of transport channel via the MAC sub-layer. It is defined in a bandwidth agnostic way, i.e., allowing it to adapt to various spectrum allocations.

The physical channels defined in the downlink are:

- Physical Downlink Shared Channel (PDSCH),
- Physical Downlink Control Channel (PDCCH),
- Common Control Physical Channel (CCPCH).

The physical channels defined in the uplink are:

- Physical Uplink Shared Channel (PUSCH),
- Physical Uplink Control Channel (PUCCH).

In addition, signals are defined as reference signals, primary and secondary synchronization signals or random access preambles. The modulation schemes supported in the downlink are QPSK, 16QAM and 64QAM, and in the uplink QPSK, 16QAM and 64QAM. The Broadcast channel uses only QPSK.

The channel coding scheme for transport blocks in LTE is Turbo Coding with a coding rate of R=1/3, two 8-state constituent encoders and a contention-free quadratic permutation polynomial (QPP) turbo code internal interleaver. Trellis termination is used for the turbo coding. Before the turbo coding, transport blocks are segmented into byte aligned segments with a maximum information block size of 6144 bits. Error detection is supported by the use of 24 bit CRC.

The generic radio frame structure as shown in Figure 2 is valid for FDD and TDD. Each radio frame consists of 20 slots of length 0.5ms, numbered from 0 to 19. A sub-frame is defined as two consecutive slots. The structures of each half-frame in a radio frame are identical.

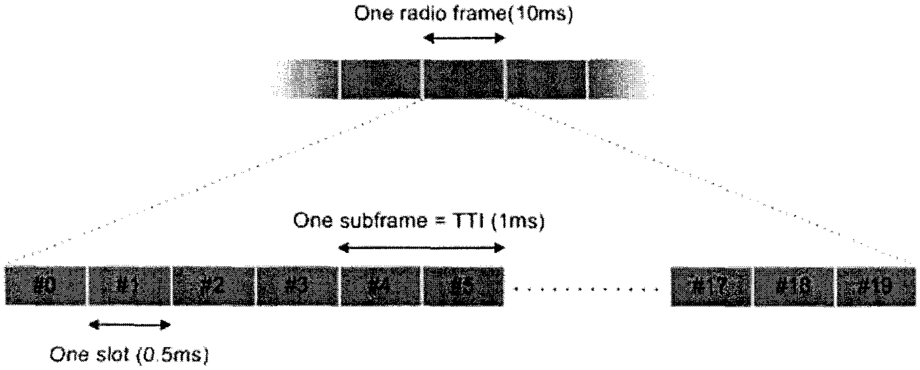

Figure 2: LTE Time-domain structure

2.3.1 Downlink OFDMA

LTE downlink transmission is based on Orthogonal Frequency Division Multiplex (OFDM). The basic LTE downlink physical resource can be seen as a time-frequency resource grid (see Figure 3), where each resource element corresponds to one OFDM sub-carrier during one OFDM symbol interval.

OFDMA distributes sub-carriers to different users at the same time, so that multiple users can be scheduled to receive data simultaneously. Sub-carriers are allocated in contiguous groups for simplicity and to reduce the overhead of indicating which sub-carriers have been allocated to each user.

OFDMA can also be used in combination with Time Division Multiple Access (TDMA), such that the resources are partitioned in the time-frequency plane, i.e., groups of sub-carriers for specific

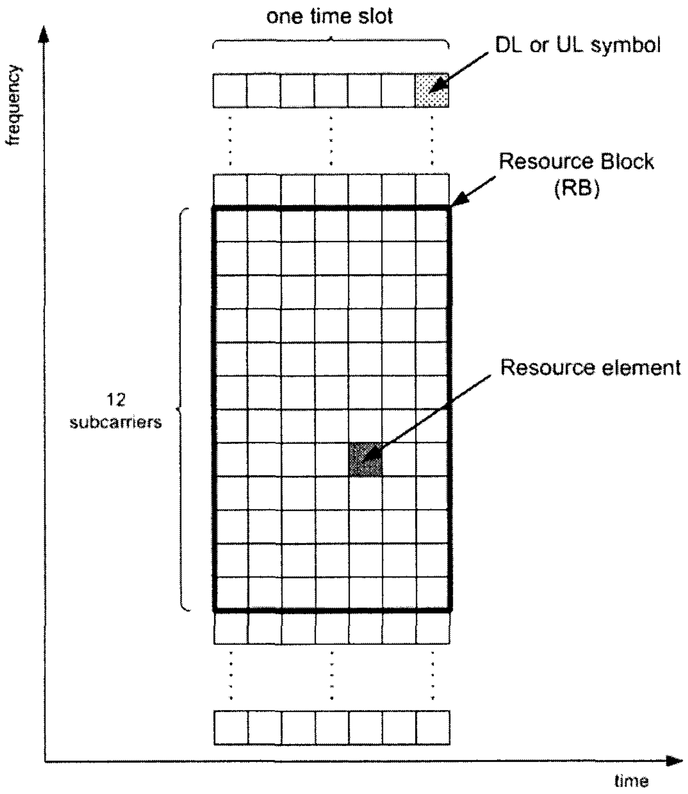

Figure 3: Resource grid

time duration. In LTE, such time-frequency blocks are known as Resource Blocks (RB). Figure 4 depicts such an OFDMA/TDMA mixed strategy used in LTE.

The practical implementation of an OFDMA system is based on digital technology and more specifically on the use of Discrete Fourier Transform (DFT) and the inverse operation (IDFT) to move between time and frequency domain representation.

For the LTE downlink, the OFDM sub-carrier spacing has been chosen to $f = 15kHz$. Assuming an FFT-based transmitter/receiver implementation, some argument for adopting a $15kHz$ sub-carrier for LTE is that it allows enough tolerance for the effects of implementation errors and Doppler effect without too much degradation in the sub-carrier orthogonality, and it simplifies the implementation of WCDMA/HSPA/LTE multi-mode terminals.

To make an Inter-symbol Interference (ISI) free transmission, LTE places a guard interval in between the sub-carriers and their spacing. Making this guard interval long enough than the maximum expected delay spread; in the environment where the system is intended to be operated, and

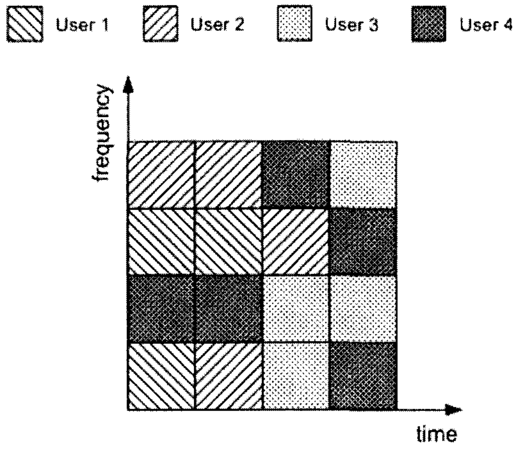

Figure 4: Example of resource allocation

considering the impact of transmitter and receiver filtering. It is possible to make a transmission completely ISI free.

2.3.2 Uplink SC-FDMA

LTE uplink transmission is based on SC-FDMA for multiple access, valid for both FDD and TDD modes of operation. SC-FDMA is a modified version of OFDMA and has similar throughput performance and almost the same complexity as OFDMA. Like OFDM, SC-FDMA also consists on sub-carriers but it transmits in sequence not in parallel as in OFDM, which prevents power fluctuations in SC-FDMA signals, i.e., low PAPR [RIK10]. Many of the requirements for LTE uplink scheme are similar to those of the downlink, but the uplink also poses some unique challenges.

- Orthogonal uplink transmission by different UE, to minimize intra-cell interference and maximize capacity.
- Flexibility to support a wide range of data rates.
- Low Peak-to-Average Power Ratio (PAPR), to reduce cost, size and power consumption of the UE Power Amplifier (PA).
- Ability to exploit the frequency diversity afforded by the wideband channel (up to 20 MHz), even when transmitting at low data rates.

- Support for frequency-selective scheduling.
- Support for advanced multiple-antenna techniques, to exploit spatial diversity and enhance uplink capacity.

Similar to the downlink, the total number of uplink sub-carriers $N_{SC} = 12N_{RB}$. In contrast to the downlink, uplink resource blocks assigned to a user equipment must always be consecutive in the frequency domain, as depicted in Figure 5. Note that, similar to the downlink, the uplink resource block is defined as 12 sub-carriers during one $0.5ms$ slot. At the same time, uplink scheduling is carried out on a sub-frame ($1ms$) basis.

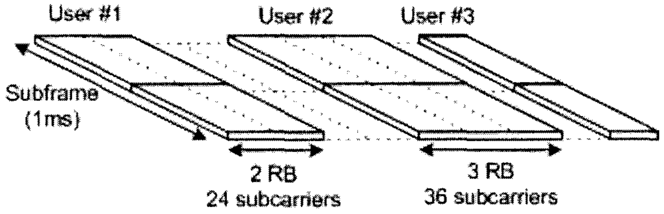

Figure 5: LTE Uplink resource allocation

2.4 Radio Interface

This section contains a brief description of the protocol layers above the physical layer. The main responsibilities of the LTE radio interface protocols are to set up, reconfigure and release the Radio Bearer.

The LTE radio interface protocol layers include Layer 2 protocols; Medium Access Control (MAC), Radio Link Control (RLC) and Packet Data Convergence Protocol (PDCP). Layer 3 consists of the Radio Resource Control (RRC) protocol, which is part of the control plane.

Figure 6 shows the overall LTE radio interface protocol architecture, covering only the protocol part of the radio access in LTE. Note that there are other protocols that are between the UE and the core network but these are transparent to the radio layers and are generally referred to as Non-Access Stratum (NAS) signaling.

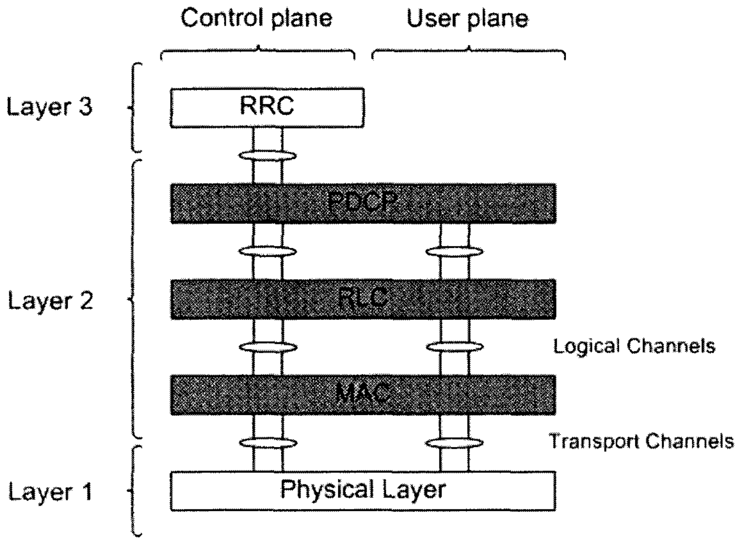

Figure 6: LTE Radio architecture

2.4.1 Medium Access Control (MAC)

The Medium Access Control layer maps the logical channels to transport channels, handles logical-channel multiplexing, hybrid-ARQ retransmissions, and uplink and downlink scheduling [3GP10c]. The connection to the physical layer is through transport channels, and the connection to the RLC layer is through logical channels. Other important tasks of the LTE MAC layer are:

- MAC layer multiplexing/demultiplexing of RLC Payload Data Units (PDU's); also Padding if a PDU is not fully filled with data.
- Traffic volume measurement reporting, to provide RRC layer information about the traffic volume experienced.
- Error correction through HARQ, to control the uplink and downlink physical layer retransmission handling in the eNodeB together with the scheduling functionality.
- Priority handling between logical channels of one UE and between UE's by means of dynamic scheduling, thus the scheduling in the eNodeB is considered as MAC layer functionality.
- Transport format selection (as part of the link adaptation functionality in the eNodeB scheduler).

2.4.2 Radio Link Control (RLC)

The Radio Link Control layer is located between the MAC layer and the PDCP layer [3GP10d]. It communicates with the PDCP layer through a Service Access Point (SAP), and with the MAC layer via logical channels. The RLC layer has the following basic functionalities:

- Transferring the PDU's received from higher layers, i.e., from RRC (Common Control Channel) or PDCP (including user plane).
- Depending on the RLC mode used, error correction with ARQ, concatenation/segmentation, in-sequence delivery and duplicate detection may be applied.
- Protocol error handling to detect and recover from the protocol error states caused by, for example, signaling errors.

The RLC can be configured in one of three data transmission modes: Transparent Mode (TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM). In AM, special functions are defined to support retransmission. When UM or AM is used, the choice between the two modes is made by the eNodeB during the RRC radio bearer setup procedure, based on the QoS requirements.

2.4.3 Packet Data Convergence Protocol (PDCP)

The Packet Data Convergence Protocol (PDCP) [3GP09] is located above the RLC layer of the user plane and PDCP is also used for most of the RRC messages. The key functionalities of the PDCP are:

- Header compression/decompression of IP packets. This is based on the Robust Header Compression (ROHC) protocol, specified in the Internet Engineering Task Force (IETF) [49]. Header compression is more important for smaller IP packets in question, especially in connection with the Voice over IP (VoIP) service, as the large IP header could be a significant source of overhead for small data rates.
- Ciphering/deciphering both the user plane and most of the control plane data.

- Integrity protection and verification, to ensure that control information is coming from the correct source.

Besides the functionalities listed above, the PDCP layer has specific functions in connection with the handover events (intra-LTE). The PDCP does the in-order delivery function in the downlink direction and detects duplicates. In the uplink direction, PDCP retransmits all the packets which have not been indicated by lower layers to be completed, as the lower layers will flush all the HARQ buffers with handover. In the downlink direction, the PDCP layer will forward the non-delivered packets to the new eNodeB. This is to ensure that no data are lost in connection with a handover event between LTE eNodeB's.

2.4.4 Radio Resource Control (RRC)

Radio Resource Control messages are a major part of the control information exchanged between the UE and E-UTRAN. The RRC protocol supports the transfer of common Non-Access Stratum (NAS) information as well as dedicated NAS information (which is applicable only to a specific UE) [3GP10e]. In addition, for UE's in RRC_IDLE, RRC supports notification of incoming calls. The RRC protocol covers a number of functional areas.

- **System information** handles the broadcasting of system information, which includes NAS common information. Some of the system information is applicable only for UE's in RRC_IDLE while other system information is also applicable for UE's in RRC_CONNECTED.
- **RRC connection control** covers all procedures related to the establishment, modification and release of an RRC connection, including paging, initial security activation, establishment of Signaling Radio Bearers and of radio bearers carrying user data, handover within LTE, configuration of the lower protocol layers, access class barring and radio link failure.
- **Network controlled inter-RAT mobility** includes (besides the mobility procedures) security activation and transfer of UE RRC context information.

- **Measurement configuration and reporting** for intra-frequency, inter-frequency and inter-RAT mobility, includes configuration and activation of measurement gaps.

2.5 Radio Resource Management

Radio Resource Management (RRM) aims to ensure that the radio resources are efficiently used, taking advantage of the available adaptation techniques, and to serve the users according to their configured QoS parameters.

Figure 7 shows an overview of the corresponding mapping of the primary RRM related algorithms to the different layers. The family of RRM algorithms at the eNodeB exploits various functionalities from Layer 1 to Layer 3.

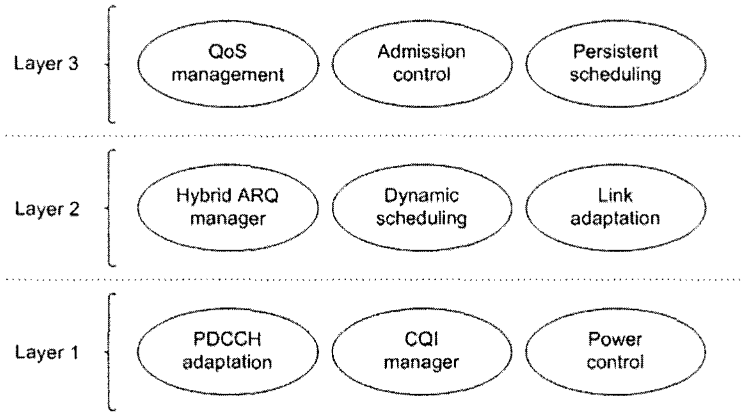

Figure 7: RRM Functions

2.5.1 Scheduling

Dynamic packet scheduling and link adaptation are key features to ensure high spectral efficiency while providing the required QoS in the cell.

The scheduler performs scheduling decisions every Time Transmission Interval (TTI) by allocating resource blocks to the users, as well as transmission parameters including modulation and coding scheme. The latter is referred to as link adaptation. In this thesis, we will refer indistinguishably to scheduling and resource allocation as our algorithms performs both activities at the same time.

There are some differences between uplink and downlink schedulers. The main differences are listed below.

- eNodeB does not have full knowledge of the amount of buffered data at the UE.
- The uplink direction is power limited due to low UE power compared to eNodeB. This means that users cannot always be allocated a high transmission bandwidth to compensate for poor SNR conditions.
- Only consecutive resource blocks can be allocated to one user in uplink due to SC-FDMA transmission. The main difference between RB allocation in downlink and uplink stems from the uplink channel design. The single-carrier constraint in LTE uplink limits both frequency and multi-user diversity.
- The uplink is characterized by high interference variability. Interference variations from TTI to TTI in the order of 15 to 20 dB make it a hard task to estimate accurately the instantaneous uplink interference.
- An uplink grant transmitted on PDCCH in TTI n refers to uplink transmission in TTI $n+4ms$. This $4ms$ delay is due to PDCCH decoding and processing time at the UE and represents a further limitation to link adaptation and channel-aware packet scheduling in uplink.

2.5.2 Admission Control

The Admission Control (AC) algorithm decides if new requests in the cell are granted admission or not. Admission control takes into account the resource situation in the cell, the QoS requirements, as well as the priority levels, and the currently provided QoS to the active sessions in the cell. A new request is only granted admission into the system if it is estimated that QoS requirements for the new session and the in-progress sessions can be fulfilled having the same or higher priority.

Each LTE bearer has a set of associated QoS parameters. All the packets within the bearer have the same QoS treatment. It is possible to modify QoS parameters of the existing bearers

QCI	Resource Type	Priority	Delay Budget	Error Loss Rate	Example Services
1	GBR	2	100 ms	10^{-2}	Conversational voice
2	GBR	4	150 ms	10^{-3}	Conversational video(Live Streaming)
3	GBR	3	50 ms	10^{-3}	Real Time Gaming
4	GBR	5	300 ms	10^{-6}	Non-Conversational video
5	non-GBR	1	100 ms	10^{-6}	IMS signaling
6	non-GBR	6	300 ms	10^{-6}	Video(Buffered Streaming), email
7	non-GBR	7	100 ms	10^{-3}	Video(Live Streaming)
8	non-GBR	8	300 ms	10^{-6}	Browsing, file download
9	non-GBR	9	300 ms	10^{-6}	file sharing

Table 1: QCI characteristics

dynamically. It is also possible to activate another parallel bearer to allow different QoS profiles for different services simultaneously.

The QoS profile of the bearer consists of the following related parameters:

- Allocation retention priority (ARP),
- Guaranteed bit rate (GBR),
- QoS class identifier (QCI).

3GPP specifications define a mapping table for nine different QCI's and their typical services [3GP10f], see Table 1. The exact admission control decision rules and algorithms are eNodeB vendor specific since they are not specified by 3GPP. In Section 3, we discuss the admission control algorithms proposed in the literature.

2.6 LTE Specific Requirements

In this section we are going to discuss the main differences from the technical specification point of view between LTE and the other 4G technology called WiMAX (Worldwide Interoperability for Microwave Access).

Although the most important similarities between these two technologies are orthogonal frequency division multiplex (OFDM) signaling, and the use of Viterbi and turbo accelerators for forward error correction. From a chip designer's perspective, if a device has to support both technologies, it means the possibility of reuse gates. From a software perspective, it's even more obvious as LTE and WiMAX share the same technology with different parameters. Flexibility, gate reuse and programmability seem to be the answers to the multi-mode LTE-WiMAX device. Here are the most important differences:

- Both technologies use orthogonal frequency division multiple access (OFDMA) in the downlink. WiMAX optimizes for maximum channel usage by organizing all the information in a wide channel. LTE, on the other hand, organizes the available spectrum into smaller blocks.

- LTE uses SC-FDMA for uplink, while WiMAX uses OFDMA in both uplink and downlink. A major problem with OFDM-based systems is their high peak-to-average power ratios. LTE adopted SC-FDMA specifically to boost Power efficiency.

  Another interesting point to mention here is that LTE (L-FDMA) adds a contiguity constraint. This constraint means that for a given user, it is only allowed to assign adjacent channels; this rule restricts the flexibility of scheduling algorithms and prevents from directly applying algorithms developed for other technologies.

- Although both standards support frequency division duplexing (FDD) and time division duplexing (TDD), WiMAX implementations are predominantly TDD. LTE is heading in the FDD direction because it is a true full-duplex operation.

Chapter 3

Literature Review

Since Scheduling and Admission Control strategies are not part of the 3GPP standard, there are many available proposals in the literature. This section provides an overview of the most relevant published work as well as an analysis of their advantages and disadvantages. We will mainly focus on LTE uplink, even though we will also cover some of the work made for WiMAX and other wireless technologies.

3.1 Scheduling

Scheduling is the process of assigning resources, so the system can provide an acceptable throughput and fairness. In this section, we will first review some basic well known scheduling algorithms, and then we will summarize a selection of scheduling strategies described in the literature.

While most of the work has been done on LTE downlink scheduling, there are already quite a few studies on the uplink scheduling. In Subsections 3.1.2 and 3.1.3, we will make a comprehensive review of the available papers in the literature, and then we will classify them as single-class or multi-class scheduling contributions.

3.1.1 Known Scheduling algorithms

1. **Round Robin**

   Round robin is one of the simplest scheduling algorithms that can be applied to scheduling problems, such as wireless packet radio networks.

   The round robin scheduler serves the users in the following way. First, the calls are placed as they arrive in the queue. The call at the beginning of the queue gets an amount of scheduled resource, and then it is placed at the end of the queue. This is repeated continuously. The amount of assigned resources will be the same for each user.

   This algorithm achieves fairness with respect to allocated resources amongst the calls. However, in general the throughput varies according to the quality of the links and therefore degrading the whole system throughput considerably, another drawback of this algorithm is that, it is not designed for multi-class systems.

2. **Proportional Fair**

   In the presence of fading, it is possible that some channels will have a higher signal to noise ratio than the average, and consequently a higher throughput. This fact can be exploited to achieve a better cell throughput using a proportional scheme.

   The scheduler can use channel quality information to calculate the capacity of the channel and prioritize users with a momentarily high available throughput. In proportional fair scheduling, the priority of each call at each resource block is calculated first, and then the user with maximum priority is assigned the RB; the algorithm continues to assign the RB to the user with next maximum priority. This process continues until all RBs are assigned or all users have been served with RB's. The priority can be calculated as the ratio of the requested data rate over the average date rate of each user.

3.1.2 Single-class Scheduling

In this subsection, we will review mainly LTE uplink approaches for single-class scheduling. Nevertheless we also include some of the designs developed for OFDMA (general approaches) and WiMAX.

In [YAAD09], the maximization of a utility function is considered to find the optimal resource allocation in Uplink SC-FDMA systems. The authors reported very good results in terms of fairness and throughput with algorithms of linear complexity.

It was also shown that the maximization of the sum throughput leads to a higher cell throughput, and that using the logarithm of throughput as a utility function ensures proportional fairness, and thus constitutes a trade-off between throughput and fairness.

One of the main disadvantages of the proposed algorithms in [YAAD09] is that the authors only work in one dimension (remember that SC-FDMA uses frequency and time) and this affects the fairness over long term intervals.

In [LPM$^+$09], the authors focused on adapting the time domain proportional fair algorithm to the LTE uplink framework, and derived from there several algorithms. In their simulations, the authors observed that maximizing the logarithm of the utility function leads to almost the same results (in terms of throughput and fairness) as maximizing the ratio between the instantaneous throughput to the total throughput.

The authors also took explicitly into account the property that requires that all the subcarriers allocated to a single user must be adjacent in frequency within each time slot. Although this contiguity constraint limits the scheduling flexibility, they incorporate this constraint while trying to maximize their own scheduling objectives. They also showed the NP-hardness of the frequency-domain scheduling problem under the block contiguity constraint.

In [MLG06], the use of the marginal utility function to formulate scheduling algorithms was proposed. Their results show that proportional fair scheduling with logarithmic user data rate can improve the sum-rate capacity.

The authors also concluded that efficient scheduling depends on accurate information about the channel response of the link between user terminals and base station. Errors in the channel estimation are mainly caused by changes in channel properties. These errors can degrade the performance of the scheduler by causing incorrect adaptation of the modulation scheme and incorrect assignment of subcarriers to users.

In [NS08], the joint subcarrier and power allocation problem with the objective of maximizing the total utility of users in the uplink of an OFDMA system were studied. In their problem formulation, the authors included the problems of sum rate maximization, proportional fairness and max-min fairness as special cases.

The authors claimed to provide a Pareto optimal solution within a large neighborhood, and an efficiently computable upper bound of the optimal solution. One important thing to mention here is that this algorithm cannot be directly applied to LTE uplink because it does not include the block contiguity constraint.

Other OFDMA uplink scheduling algorithms were devised in [GC08]. Again, the results were analyzed in terms of throughput and fairness. The authors reported that the reduction in complexity and fairness is achieved by performing scheduling in two steps: firstly, they allocate subcarriers by considering per-user fairness; and secondly, they apply residual subcarrier allocation to increase the sum rate.

In addition to the heuristics, few authors attempted to solve exactly the scheduling and subcarrier allocation problem using an optimization model, see [HSAB09] [KHK05] [GC08]. They obtain near optimal solutions that usually outperform the heuristic solutions. However, none of those studies include subcarriers or block contiguity constraints.

The major drawback of all these proposals is that they only consider throughput in the development of their algorithms. Although throughput is a very important factor, it is not the only one, and in fact as we will show in the next chapters, it is very important to incorporate more parameters.

3.1.3 Multi-class Scheduling

In this subsection, we will review LTE uplink approaches for multi-class scheduling and some of the designs developed for OFDMA (general approaches).

In [QRTC09], Qian *et al.* developed a resource block allocation algorithm for multi-service downlink LTE systems and deduced from it a scheduling algorithm that takes advantage of the queue state information in the data-link layer and the channel state information in the physical layer. Results were analyzed in terms of average user throughput and number of satisfied users.

One of the key features of this paper is that all users are classified into three categories and the scheduler assigns resource blocks in turns. An important disadvantage is that they defined data loss rate as the only QoS indicator.

In [PKF$^+$09], Petersen *et al.* showed a comprehensive study of the many resource allocation techniques available for LTE downlink. The quality of service is outlined as being one of the main objectives to maximize the system capacity while serving all users according to their minimum QoS requirements. It is also shown how the radio resource management algorithms at the base station offers opportunities for efficient designs due to the easy access to air interface measurements.

In [WWS07], the authors investigated how call setup signaling is affected as the load on the system increases in a traffic scenario where all users engage in both a voice and a video session.

They found that if all traffic is scheduled with the same priority, setup signaling might be affected and the delivery of signaling messages cannot be guaranteed. However, if these messages have higher priority, the length of call setups and terminations can be kept at almost constant values even if the system load is high. Their results also indicate that other service qualities are not significantly affected by such a priority scheme.

In [MPKM08], Monghal *et al.* introduced a decoupled time/frequency domain scheduler approach. They showed that fairness can be controlled with frequency domain metric weighting or Time Domain Priority Set Scheduling depending on the number of users in the cell.

Again in this paper, the main concern is only the throughput as a main QoS parameter.

In [ZHWW07], Liu and Xinglin proposed a simplified layered scheduling scheme based on Maximum Delay Utility for OFDM networks. In this scheme, the scheduling is divided into two main steps, macro and micro scheduling. In the macro step, the utility functions are defined, and the priorities of each type of service are determined. In the micro step, the scheduling is performed among all users inside a given traffic type determined in macro step.

The authors claim that their simplified layered scheduling has much lower computational complexity and almost the same utility values as the standard maximum delay utility scheduling while handling multiple traffic types with diverse QoS requirements.

In [KSC09], the authors proposed a scheduling algorithm that satisfies the QoS requirements of the real-time traffic and maximizes the utility of the non real-time traffic. A step-by-step approach was used to achieve these two objectives with low complexity and traffic class prioritization.

A well-known bipartite matching algorithm was adopted for the QoS scheduling of the real-time traffic and a standard gradient scheduling algorithm for the utility maximization scheduling of the non real-time traffic.

Although there are many more research work made in this topic (for FDMA, CDMA, etc) they cannot be directly applied to LTE because of its particular constraints (see Section 2.6 for details) .

3.2 Admission Control

Admission Control is the process which evaluates if the system has enough available resources to accept new incoming connections, and then decide either to accept or reject new connections so it can provide an acceptable QoS for new and existing connections.

Although there are quite a good number of studies on admission control, not many of them focus on multi-class admission control for LTE systems. In this section, we will first discuss the general approaches and then, we will focus on the available papers for LTE and WiMAX.

3.2.1 Known Admission Control algorithms

Here, we will briefly discuss three well known admission control algorithms: the Reservation Scheme, the Linear Weighting Scheme, and the Distributed Admission Control Scheme.

1. **Reservation Scheme (RS)**

   In this first scheme, the admission control algorithm reserve channels for different classes of traffic according to a certain distribution of traffic in a specific cell, i.e., if a handoff call request admission, the algorithm evaluates if the number of channels reserved for handoff allows the system to accept this request.

2. **Linear Weighting Scheme (LWS)**

   The linear weighting scheme adds to the reservation scheme that it considers the mean number of calls underway in all cells within a maximum number of hops from the originating cell in determining the admission. It uses the same principle of reserving channels for different classes of traffic but instead of just looking to one cell; it evaluates a defined group of cells and takes a decision based on the average.

3. **Distributed Admission Control Scheme (DACS)**

   Unlike the reservation scheme, in this approach, the admission is taken in a distributed manner by the periodic exchange of information among the base stations regarding the current load conditions.

   Let P_{QOS} be the highest tolerable hand-off dropping probability for all the calls of the same class. The admission controller must fulfill two critical requirements: first of all, by admitting the new call, the QoS of existing calls in the system must be maintained, and secondly the system must provide the newly admitted call with its desired QoS. Hence, a new call is admitted into the system, if the above requirements are satisfied.

3.2.2 Single-class Admission Control

In this subsection, we will review the LTE approaches for single-class systems as well as some of the strategies developed for WiMAX. Remember that WiMAX is the closest technology to LTE, as it uses a similar set of radio technologies (OFDMA).

In [ARC$^+$08], Anas *et al.* proposed an admission control algorithm for LTE uplink utilizing the fractional power control formula [3GP07b]. The algorithm determines if a user requesting admission can be accepted based on the path-gain so as to fulfill the QoS requirements of the new and existing users. The authors also claim that their algorithm tunes itself for many load conditions without additional complexity.

Unfortunately, in this paper, the authors did not address the study of QoS differentiation for mixed traffic. In addition, they did not care of the fact that in a real situation the closed loop adjustments of fractional power control should be taken into account.

In [TC08], Tarhini *et al.* proposed a new state-dependent admission control scheme where the degree of acceptance depends on the density of the users in different areas inside a single OFDMA-based WiMAX cell, users are allowed to move internally among those areas as well as externally to other cells.

The authors claim that their density-based admission control algorithm shows a lower dropping probability of on-going calls without increasing too much the blocking probability of new calls.

In [TJP08], Teh *et al.* proposed an admission control scheme that uses the bit rate statistics of each flow to predict the percentage of packets that are excessively delayed. Consequently, a new flow is allowed only if this predicted percentage is sufficiently low.

In their simulation results, they showed that their proposed method can predict the proportion of delayed packets and control the load of the network to achieve the desired delay bounds.

In [JS07], Jing and Sampalli developed an admission control scheme that handles the intra-cell mobility issue in the downlink of broadband wireless networks with link adaptation. They decomposed the cell into rings and associate resource consumption with each ring, and modeled the intra-cell mobility as a BCMP (network of interconnected queues) queuing chain network.

Here, the authors assumed that all users have the same QoS requirements. However, the authors consider that the BCMP queuing network model can be extended to a multi-class BCMP queuing network model.

Additionally, their cell mobility-based admission control algorithm can provide efficient resource allocation by predicting min-guaranteed resource consumption on a per cell basis.

In [ZXLH08], Zhu *et al.* proposed a fair connection admission control scheme (FCAC) to improve the system fairness on bandwidth allocation to users by setting a dynamic admission threshold. FCAC considers a common scenario in which users are limited to the fixed amount of bandwidth by Internet Service Provider.

In this solution, the users required bandwidth and the available system bandwidth are used in the admission control criterion. In order to improve the system fairness on bandwidth allocation to users, the bandwidth requirements of other users are also taken into account to decide whether to admit a new user connection request.

3.2.3 Multi-class Admission Control

Multi-class admission control is a key factor to provide new services. As it is well known, new mobile technologies aim to provide multiple services, i.e., voice, video, internet, over the same network, and consequently each type of traffic has its own requirements. In this subsection, we will review the LTE and WiMAX admission control approaches for multi-class systems.

In [BCC$^+$09], Bae *et al.* proposed a Delay-aware Admission Control algorithm (DACAC) that provides QoS for various kinds of services in LTE system. Their algorithm utilizes statistical data for packet delay and Resource Block utilization in order to guarantee packet delay requirements for on-going calls.

They also report that DACAC algorithm yields lower hand-off call dropping probability than new call blocking probability, leading to a support of continuous communications for users moving between different cells.

In [KMS08], an Admission Control scheme for Mobile WiMAX that considers both bandwidth and delay for the admitted connections was proposed.

This admission control scheme considers a priority scheme for different types of traffic, for example provides higher priority to hand-off connections, than to a newly incoming connection.

In [RQL07], Rong *et al.* addressed the admission control problem from the perspectives of WiMAX service providers and subscribers. They formulate the admission control as an optimization problem, in which the demands of service providers and subscribers are taken into account.

In addition, the authors developed a computational method for their strategy, and a utility and fairness constrained greedy revenue algorithm considering many classes of traffic. The authors argue that its proposed admission control approach can meet the requirements of both service providers and subscribers.

In [SCK10], Saddoud *et al.* proposed an admission control scheme (sometimes called bandwidth allocation in this paper) based on QoS requirements of different traffic types for mobile WiMAX.

Their admission control algorithm gives priority to real time service classes minimizing in this way the call blocking probability for different traffic types.

3.3 Joint Scheduling and Admission Control

In this section, we will review the LTE approaches that combines scheduling and admission control for multi-class systems. We also include some of the general approaches that can be applicable to LTE.

In [QHS$^+$09], Qian *et al.* proposed an admission control scheme for handling multiclass services. To solve the optimization problem, they presented a combined complete sharing and virtual partitioning resource allocation model and develop a service degradation scheme in case of resource limitations in their proposed admission control scheme.

Their proposal is based on identifying an optimal proportion for different service groups that maximize the system capacity while maintaining the QoS of all admitted users.

In [LYZ$^+$08], Lei *et al.* proposed a resource allocation algorithm and an admission control scheme for LTE systems with heterogeneous services. Their proposed allocation algorithm introduces a transmission guard interval which gives high priority to real-time service approaching their delay limit, and an admission control algorithm that can adaptively adjust the threshold according to the network condition.

The authors reported that their proposed AC scheme can balance ongoing connections of different classes of traffic and it is potentially easy to reserve resources to support handover users.

In [MA08], Anas *et al.* developed a combined admission control and a decoupled time-frequency domain scheduling for LTE uplink that is able to differentiate between user classes, for a realistic mix of committed bit rate and best effort traffic.

The authors report that their admission control algorithm blocks the bearer with a very low path-gain, and fulfill the required GBR of admitted bearers with a low outage probability.

One important thing missing in this study is to focus on the scheduling algorithms to differentiate and prioritize the GBR bearers over non-GBR bearers.

In [Hos03], Hosein presented a flexible framework for admission control and scheduling within which a wireless operator can provide QoS services, and at the same time, gives users priority based on their subscription plans. This framework allows the operator to fine tune their pricing scheme in order to maximize revenue.

The admission control approach addressed here consist on imposing penalties if a connection is accepted and the user is outside the limits defined by a barrier function and enforced by the scheduler. The QoS metric used here is only based on throughput measurements, as the author considers that it is the most resource intensive parameter.

In [TTLK08], Tung *et al.* proposed an adjustable Quadra-threshold (QTBR) dynamic call admission control scheme and QoS aware bandwidth allocation algorithm for WiMAX. Their proposed scheme supports voice, data and multimedia services with differentiated QoS.

These schemes works together to keep the resource utilization in the subscriber station close to 100%. Additionally, the authors report that the QoS aware bandwidth allocation with threshold setting lowers the Blocking Probability of real time and non real time traffic.

In [TLT$^+$07], Tsang *et al.* proposed a QoS solution incorporating a dynamic call admission control scheme and a bandwidth allocation algorithm for WIMAX mobile application. For the admission control, they use an adjustable Quadra-threshold (QTBR) scheme and a QoS aware scheme for the resource allocation.

The authors claimed that their proposed solution supports voice, data and multimedia services with differentiated QoS, and promise to maintain the resource utilization close to the limit while controlling the dropping probability.

In [CG07], Chaudhry and Guha proposed an admission control scheme and a scheduling algorithm for fixed and mobile Broadband Wireless Access systems. The admission control scheme reserves an adaptive temporal channel bandwidth for mobile subscriber stations based on the most recent requests to assure seamless hand-off, while the scheduler allocates physical layer slots to user packets based on the application's data rate and latency characteristics.

Additionally, their proposed scheduling algorithm prioritizes real-time over non real-time traffic in accordance with the Quality of Service (QoS) parameters of service flows defined in the standard.

Chapter 4

Proposed Mathematical Model and Algorithms

In this chapter, we expose the mathematical model developed for a single cell network, based on the technology and constraints discussed in the previous chapters. Then we move to heuristic approaches, the first approach it is to consider scheduling for a single-class system. Then, we discuss how to extend the approach to a multi-class system, and finally we propose a joint solution that guarantees admission control and scheduling with QoS and GoS in a multi-class environment. Simulation results that verify the theoretical analysis will be discussed in the next chapter.

4.1 Single-class Scheduling

In this section, we propose an original optimization model for the resource allocation problem in the context of a Single-Cell SCFDMA (L-FDMA scheme) subject to delay constraints. It corresponds to a constrained optimization problem in which the objective is to maximize a utility function that captures the fairness concern.

We consider a Single Cell SC-FDMA transmission system with a bandwidth of B Hz. In time domain, the basic unit of scheduling is a transmit time interval (TTI) where typically 1 TTI = 1ms.

The total bandwidth is partitioned in Resource Blocks (RB's) and each RB has L subcarriers, one (or more) RB can be assigned to a given user in each TTI. We assume that the base station has a perfect knowledge of the channel gains of all users in the frequency and time domains.

4.1.1 LTE Resource Allocation Model with Delay Requirement

Although some mathematical models have already been proposed for the scheduling and resource allocation in LTE systems, none of them has yet expressed the block contiguity constraints together with the delay constraints. Among the previous references, we can mention those of [HSAB09] [GC08] [KHK05] where the authors deal with scheduling and subcarrier allocation without any subcarrier contiguity or block constraints. The model that is described below is written assuming we manipulate resource blocks (1 RB = 12 subcarriers).

We now describe the mathematical model starting by its objective. In economics, utility is a concept representing the level of satisfaction from consumption of goods and services, and it is used for balancing the efficiency and fairness among users. In wireless communications, several concerns have already been expressed through a utility function, among which power, throughput, or user fairness. Since user data rate is an important parameter to determine user satisfaction and fairness in wireless communications, a utility function defined as an increasing function of user data rate is often selected. We will remain with such a choice and therefore propose a first objective aiming at maximizing the sum of the user utility at each TTI. Alternate objectives will be discussed at the end of this section.

Let N be the set of blocks and K be the set of users. For a given user $k \in K$, we denote by N_k the set of resource blocks allocated to user k, by P_k the instantaneous transmission power of user k, by $P_k^{\max}$ the maximum transmit power, and by R_k its achievable throughput.

The maximization of the user utilities can be written as follows:

$$\max \sum_{k \in K} U(R_k | N_k) \tag{1}$$

where $U(R_k|N_k)$ is the utility of user k as a function of the throughput R_k given the allocation of RBs $n \in N_k$ to user k.

The block resource allocation corresponds to the following set of allocations:

$$\sum_{k \in K} \alpha_{n,k} = 1 \qquad n \in N \tag{2}$$

$$\alpha_{n,k} \geq 0 \qquad n \in N; k \in K \tag{3}$$

$$0 \leq P_k \leq P_k^{\max} \qquad k \in K \tag{4}$$

where $\alpha_{n,k} = 1$ if RB n is allocated to user k. Constraints (2) express the condition that each RB can be allocated only to a single user during one TTI. Constraint (3) is a non negativity constraint, and constraints (4) convey the power limitations of each user,

In order to guarantee the resource block contiguity constraints, we need to introduce two new set of variables $\delta^+_{n,k}$ and $\delta^-_{n,k}$ such that:

$$\alpha_{n+1,k} = \alpha_{n,k} + \delta^+_{n,k} - \delta^-_{n,k} \qquad n \in N, \; k \in K \tag{5}$$

$$\delta^+_{n,k} + \delta^-_{n,k} \leq 1 \qquad n \in N \tag{6}$$

$$\sum_{n \in N} \delta^+_{n,k} \leq 1 \qquad k \in K \tag{7}$$

$$\sum_{n \in N} \delta^-_{n,k} \leq 1 \qquad k \in K \tag{8}$$

$$\sum_{k \in K} \delta^+_{n,k} = 1; \sum_{k \in K} \delta^-_{n,k} = 1 \qquad n \in N \tag{9}$$

$$\delta^+_{n,k}, \delta^-_{n,k} \in \{0, 1\} \qquad n \in N \tag{10}$$

Constraints (5) detect whether there is an increase or a decrease of the value of $\alpha_{n,k}$ when compared to $\alpha_{n+1,k}$ taking into account that, due to constraint (6), there cannot be a decrease at the same time than an increase. Constraint (7) does not allow more than one increase, and therefore guarantees that once a block is selected while the previous one was not, all subsequent resource blocks are either all selected, or once one is not selected, the subsequent resource blocks of this latter block are not selected. Constraint (8) does not allow more than one decrease, and therefore guarantees that as soon as a RB is not selected while the previous one was, all subsequent RBs cannot be selected. Note that constraints (8) are redundant, once constraints (7) are satisfied, constraints (8) are automatically satisfied as well. Constraints (9) ensure, similarly to (2), that each user is assigned to at most one RB.

Note that it would be easy to add constraints setting limits on the size of sequences made of contiguous RBs with the addition of integer variables $x_{n,k}$ such that $x^+_{n,k} = n\delta^+_{n,k}$ and $x^-_{n,k} = n\delta^-_{n,k}$: length of the sequence of contiguous RBs for is user k is then $\sum\limits_{n \in N} x^+_{n,k} - \sum\limits_{n \in N} x^-_{n,k}$.

Last, we need to express the delay constraints. Firstly, we express that each user cannot exceed the maximum allowable delay and secondly, that its throughput cannot be smaller than the minimum required throughput for each user.

$$\mathrm{d}_k \leq \mathrm{d}^{\max} \qquad k \in K \tag{11}$$

$$R_k \geq R^{\min} \qquad k \in K, \tag{12}$$

where $\mathrm{d}^{\max}$ is the maximum allowed delay and $R^{\min}$ the minimum acceptable throughput for user k within the LTE cell.

It is important to mention that, we assume that we always have at least one user in the system, and we only deal with resource block assignment. It might be possible that, a given user which has been assigned resources has not information to transmit in a specific time slot.

4.1.2 Alternate Utility Functions

As mentioned earlier, the utility function is often defined as dependent of the throughput. However, if we assume that the utility function is equal to the sum of the throughputs; users located at the edge of the cell will always be allocated fewer resources than the users near the base station. This is due to the path loss (proportional to the distance). To address this issue, the sum of the utilities is often replaced by the sum of the logarithmic utility functions in order to provide a proportional fairness. Hence, we can either use

$$\max \sum_{k \in K} U_k(R_k|N_k) = \max \sum_{k \in K} \sum_{n \in N} \alpha_{n,k} R_k \tag{13}$$

or

$$\max \sum_{k \in K} \ln U_k(R_k|N_k) = \max \sum_{k \in K} \sum_{n \in N} \alpha_{n,k} \ln(R_k) \tag{14}$$

where ln represents the natural logarithm, and $N_k = \{n \in N : \alpha_{n,k} = 1\}$. Note that not all users are necessarily scheduled if not enough resource blocks are available, or if the number of available resource blocks vs. the number of users do not allow the satisfaction of all the quality of service (QoS) requirements for all users.

Another way to maximize this proportional fair objective is to use:

$$\lambda_{n,k} = r_{n,k}/R_k \tag{15}$$

where $r_{n,k}$ is the instantaneous data rate for user k on RB n (see [AQS05], [KW04]). We can establish another version of (13) or (14) as follows:

$$\max \sum_{k \in K} \sum_{n \in N} \alpha_{n,k} \lambda_{n,k} \tag{16}$$

where $\lambda_{n,k}$ is defined as in (15).

Seeking to maximize the general objective in (1) or (16) under the delay, throughput, and block contiguity constraints leads therefore to a difficult mathematical programming problem, due mainly to the nonlinear and non-convex objective. In spite of the evolution of large scale 0-1 programming techniques, it is difficult to hope that they would scale in the context of real-time scheduling and resource allocation Uplink LTE systems, although they could be useful in the context of either developing accurate planning tools, or assessing the quality of the heuristic solutions. In the next Section, we focus on the design of two efficient greedy heuristics, and, in the following section, on their performance evaluation.

4.1.3 Scheduling and Resource Allocation Algorithms

In this section, we propose two new algorithms that perform scheduling and subcarrier allocation within LTE Uplink system model.

Algorithm1: LC-Delay

The first greedy algorithm, called Algorithm 1 LC-delay, schedules each resource block RB to a user k in a way that maximizes the marginal utility, satisfies the maximum allowed delay, and guarantees a minimum throughput $R^{\min}$ for each user.

Algorithm 1 LC-Delay

```
{Initialization}
$U_{RB} \leftarrow$ index set of unassigned resource blocks
$I_{\mathrm{RB},k} \leftarrow$ set of assigned RB's to user $k$.
At the outset, $I_{\mathrm{RB},k} \leftarrow \emptyset$.

{Main iteration}
for each $\mathrm{RB}_n$, $n \in U_{RB}$ do
  if $\mathrm{RB}_n$ has not been assigned a user then
    if $\max_{k \in K} \mathrm{d}_k \leq \mathrm{d}^{\max}$ and $\min_{k \in K} R_k \geq R^{\min}$ then
      $K' \leftarrow K$; AssignRB $\leftarrow$ .FALSE.
      while AssignRB = .FALSE. and $K' \neq \emptyset$ do
        $k^\star \leftarrow$ user of $K'$ with largest marginal utility
        if $\mathrm{RB}_n$ is adjacent to the resource blocks in $I_{\mathrm{RB},k^\star}$ or $I_{\mathrm{RB},k^\star} = \emptyset$ then
          Assign $\mathrm{RB}_n$ to $k^\star$ ; AssignRB = .TRUE.
        else
          $K' \leftarrow K' \setminus \{k^\star\}$
        end if
      end while
    else
      $K' \leftarrow K$; AssignRB $\leftarrow$ .FALSE. ; $x \leftarrow 1$
      while AssignRB = .FALSE. and $K' \neq \emptyset$ do
        $k^\star \leftarrow$ user of $K'$ with largest delay or smallest throughput
        Select RB$m$ such that it is associated with the $x$th largest marginal utility
        if RB$m$ is adjacent to the resource blocks in $I_{\mathrm{RB},k^\star}$ or $I_{\mathrm{RB},k^\star} = \emptyset$ then
          Assign RB$m$ to $k^\star$ ; AssignRB = .TRUE.
        else
          $x \leftarrow x + 1$ ; $K' \leftarrow K' \setminus \{k^\star\}$
        end if
      end while
      if $\mathrm{RB}_n$ has not been assigned a user then
        $K' \leftarrow \{k \in K : \mathrm{d}_k \leq \mathrm{d}^{\max}; R_k \geq R^{\min}\}$; AssignRB $\leftarrow$ .FALSE.
        while AssignRB = .FALSE. and $K' \neq \emptyset$ do
          $k^\star \leftarrow$ user of $K'$ with largest marginal utility
          if $\mathrm{RB}_n$ is adjacent to the resource blocks in $I_{\mathrm{RB},k^\star}$ or $I_{\mathrm{RB},k^\star} = \emptyset$ then
            Assign RB$n$ to $k^\star$ ; AssignRB = .TRUE.
          else
            $K' \leftarrow K' \setminus \{k^\star\}$
          end if
        end while
      end if
    end if
    Remove from $U_{RB}$ all the blocks which have been assigned a user
  end if
end for
```

The marginal utility is defined as follows: Let N be the set of resource blocks. For resource block RB_n

$$\Lambda_{n,k} = U(R_k | I_{\mathrm{RB},k} \cup \{n\}) - U(R_k | I_{\mathrm{RB},k}). \tag{17}$$

Assuming

$$N_k = \{n \in N : n \in I_{\mathrm{RB},k}\}, \tag{18}$$

The marginal utility can be rewritten:

$$\Lambda_{n,k} = U(R_k | N_k \cup \{n \in N\}) - U(R_k | N_k). \tag{19}$$

The LC-Delay algorithm goes through each block, one after the other one, and assigns it to a user, taking into account the adjacency resource block constraints and if the maximum delay and minimum throughput requirements are satisfied for all users. Otherwise, it assigns RB's, first to the users with critical delay or throughput; as long as resource blocks are available and adjacent resource block constraints are satisfied. Note that some users may never be assigned if their marginal utility is always much smaller than the ones of the other users (in particular when the number of users is larger than the number of resource blocks).

Algorithm2: PF-Delay

The second proposed algorithm consists of allocating a resource block RBn that maximizes the metric value λ_k^n in a way that users will never experience a delay greater than the maximum allowed delay (if they have a reasonably high utility value).

PF-Delay algorithm differs from the LC-Delay one in the use of a different metric to assign channels: Instead of using the marginal utility, we use the proportion between the current throughput to the total throughput. In addition, we do not assign the resource blocks in order, but with respect to the users with the most critical delay requirement, under the condition that the user has a reasonable utility value. Indeed, again, as in the LC-Delay algorithm, some users may never be assigned if their marginal utility is always much smaller than the ones of the other users.

Algorithm 2 PF-Delay

```
{Initialization}
Assumption: $\lambda_{n,k} = r_{n,k}/R_k$ PF metric values are sorted in the decreasing order of their value.
$U_{RB} \leftarrow$ set of unassigned resource blocks
$I_{\mathrm{RB},k} \leftarrow$ set of assigned RB's to user $k$
At the outset, $U_{RB} \leftarrow \emptyset$ ; $I_{\mathrm{RB},k} \leftarrow \emptyset$
$x \leftarrow 1$

while $U_{RB}$ is not empty do
  if $\max_{k \in K} \mathrm{d}_k \leq \mathrm{d}^{\max}$ then
    Select RB$n$ such that it is associated with the $x$th largest PF metric value
    Let $k^\star$ be the corresponding user
    if RB$_n$ is adjacent to the resource blocks in $I_{\mathrm{RB},k^\star}$ or $I_{\mathrm{RB},k^\star} = \emptyset$ then
      Assign RB$_n$ to user $k^\star$ ; $U_{RB} \leftarrow U_{RB} \setminus \{\mathrm{RB}_n\}$
      $x \leftarrow 1$
    else
      $x \leftarrow x + 1$
    end if
  else
    Assign best available channel to the user with the largest delay
    Remove from $U_{RB}$ all the blocks which have been assigned a user
  end if
end while
```

4.2 Multi-class Scheduling

In this section, we define the resource allocation problem as a constrained utility maximization problem in the context of a Single-Cell SC-FDMA (L-FDMA scheme) subject to delay and rate constraints, in the context of a traffic with different class of services.

4.2.1 Resource Allocation Model

Let N be the set of resource blocks and K be the set of requests. We assume that there are different classes of services. Let I be the index set of classes, with generic index i. For a given request $k \in K$, it belongs to a given class C_i of service with specific delay and rate requirements, i.e., a maximum delay $\mathrm{d}_i^{\max}$ and a minimum rate (throughput) $R_i^{\min}$ for each class C_i of service. For a given request $k \in K$, we further denote by N_k the set of resource blocks allocated to request k, by P_k the instantaneous transmission power of request k, by $P_k^{\max}$ the maximum transmission power, by R_k its achievable throughput, and by C_k the class of service. In addition, depending on the class of services a given request belongs to, there are different delay and rate requirements, As in the previous section, the mathematical model can be expressed as in equations (1) to (10). We express the constraint that each request cannot exceed the maximum allowable delay and that its throughput cannot be smaller than the minimum throughput allowed for each request, depending on its class of service:

$$\mathrm{d}_k \leq \mathrm{d}_i^{\max} \qquad k \in K_i, i \in I \tag{20}$$

$$R_k \geq R_i^{\min} \qquad k \in K_i, i \in I, \tag{21}$$

where $\mathrm{d}_i^{\max}$ is the maximum allowed delay of class i and $R_i^{\min}$ the minimum acceptable throughput of class C_i for request k within the LTE cell.

To provide a proportional fairness the sum of the utilities is often replace by the sum of the logarithmic utility functions. Hence, we can either use

$$\max \sum_{k \in K} U_k(R_k | N_k) = \max \sum_{k \in K} \sum_{n \in N} \alpha_{n,k} R_k \tag{22}$$

or

$$\max \sum_{k \in K} \ln U_k(R_k | N_k) = \max \sum_{k \in K} \sum_{n \in N} \alpha_{n,k} \ln(R_k) \tag{23}$$

where ln represents the natural logarithm, and $N_k = \{n \in N : \alpha_{n,k} = 1\}$. Note that not all users are necessarily scheduled if not enough resource blocks are available, or if the number of available resource blocks vs. the number of users does not allow the satisfaction of all the quality of service (QoS) requirements for all users.

Another way to maximize this proportional fair objective (see [AQS05], [KW04]) is to use:

$$\max \sum_{k \in K} \sum_{n \in N} \alpha_{n,k} f_k \tag{24}$$

where f_k is a proportion defined as follows

$$f_k = \frac{R_k \times \mathrm{d}_i^{\max}}{R_i^{\min} \times \overline{\mathrm{d}_i}} \tag{25}$$

and $\overline{\mathrm{d}_i}$ is the average delay of class i:

$$\overline{\mathrm{d}_i} = \sum_{k \in \text{ class } i} \mathrm{d}_k.$$

Although seeking to maximize the general objective in (23) or (24) is equivalent, we will use the approach as defined in (24) since when using the proportional fair scheduling metric f instead of the typical proportional fair metric allows giving more priority to the delay requirement to requests that are closer to the maximum delay or to the minimum throughput requirement. Another point of interest is that using such an objective leads to a much more scalable ILP problem. However, in the context of real time requirements, it might not be scalable enough; this is the reason why we focus on heuristics in the next section.

4.2.2 Scheduling and Resource Allocation Algorithms

In this section, we propose two new scheduling algorithms for LTE Uplink systems, called Single channel scheduling algorithm (SC-SA) and multiple channel scheduling algorithm (MC-SA) in the context of multiple class traffic. Consequently, both heuristics guarantee the throughput and delay requirements of each request according to its own QoS characteristics (multiple classes of traffic). While the SC-SA heuristic allows the assignment of at most one channel block per request and per TTI, independently of its QoS requirements, the second heuristic allows the assignment of more than one channel block per request.

The packet scheduling is done as a one step algorithm that combines time-domain (TD) scheduling and frequency-domain (FD) scheduling by selecting the request which will be multiplexed in time and frequency with the same metric, as explained in the next sub-sections.

When the number of requests is smaller than the maximum number of resource blocks, both algorithms equally distributes the available RB's among the requests in the cell. This is not necessarily the fairest way to distribute the resource blocks. However, there is no necessity to devise a complex resource block share among the requests in such a case.

Single channel scheduling algorithm (SC-SA)

The SC-SA algorithm dynamically adapts each request throughput to the cell load while trying to meet the multiclass QoS requirements.

Algorithm 3 SC-SA

```
{Initialization}
$U_{RB} \leftarrow$ set of unassigned resource blocks
Define $F = \{f_k \times \delta_k^{\mathrm{d}} \times \delta_k^R : k \in K\}$
Sort the values of $F$ in ascending order.
for all $k \in K$ do
    $\delta_k^{\mathrm{d}} = 0$ if $\mathrm{d}_k \geq \mathrm{d}_i^{\max}$ and $k \in C_i$, 1 otherwise
    $\delta_k^R = 0$ if $R_k \leq R_i^{\min}$, and $k \in C_i$, 1 otherwise
end for
$x \leftarrow 1$
while $U_{RB} \neq \emptyset$ do
    if $|K| < RB$ then
        Let $m = \lfloor RB/|K| \rfloor$
        Assign the channels in order so that $m$ adjacent channels are assigned to the last $|K| - 1$ requests in $F$
        Assign the remaining adjacent channels to the first request in $F$
    else
        Set $k^*$ to the index of the $x$th value of $F$
        Assign $RB_x$ to request $k^*$
    end if
    $U_{RB} \leftarrow U_{RB} \setminus \{RB's$ assigned in previous step$\}$
    $x \leftarrow x + 1$
end while
```

Algorithm 3 SC-SA describes in detail the allocation scheme. After the initialization phase, in which the scheduler captures the necessary information from the UE (User Equipment), the value of the metric f_k is calculated for every request $k \in K$. During the first iterations, the scheduler assigns RB's in sequential order as long as it does not have accurate information on the delay or rate performances for the requests (we indeed start from scratch in our simulation) After a few iterations, there is enough information in the system to execute the algorithm with the QoS concerns.

In case the number of requests is smaller than the total number of available RB's, the scheduler distributes proportionally the remaining RB's among all the requests. If the number of requests is higher than the total number of available RB's, the scheduler takes the requests that are experiencing the poorest performance and assign them a RB. Finally, once the allocation is performed, the system updates all the relevant parameters.

Note that the system adjusts itself in order to match the QoS target. The aim of the SC-SA allocation scheme is to allocate more RB's to requests experiencing the worst (most stringent QoS) conditions, than the ones which have a higher performance. This results in an increased system capacity. The underlying assumption in the SC-SA algorithm is that the channel, in a steady-state, will have a similar behavior over consecutive RB's.

Multiple channel scheduling algorithm (MC-SA)

The MC-SA algorithm consists of allocating more than one resource block to the requests that are not meeting the throughput target. The reason for allocating multiple RBs to a single request is in order to help the requests such that the average throughput transmitted within one TTI is smaller than the throughput target. Note that a block being made of 12 subcarriers, the delay requirement can always be met using only one single resource block during one TTI.

MC-SA uses the same metric value f_k to assign RBs as the SC-SA algorithm for ordering the requests according to their QoS requirements, but a different strategy for assigning channels, and in particular for deciding how many channels should be assigned.

Algorithm 4 MC-SA

```
{Initialization}
$U_{RB} \leftarrow$ set of unassigned resource blocks
Define $F = \{f_k \times \delta_k^{\mathrm{d}} \times \delta_k^R : k \in K\}$
Sort the values of $F$ in ascending order.
for all $k \in K$ do
    $\delta_k^{\mathrm{d}} = 0$ if $\mathrm{d}_k \geq \mathrm{d}_i^{\max}$ and $k \in C_i$, 1 otherwise
    $\delta_k^R = 0$ if $R_k \leq R_i^{\min}$, and $k \in C_i$, 1 otherwise
end for
$x \leftarrow 1$
while $U_{RB} \neq \emptyset$ do
    if $|K| < RB$ then
        Let $m = \lfloor RB/|K| \rfloor$
        Assign the channels in order so that $m$ adjacent channels are assigned to the last $|K| - 1$
        requests in $F$
        Assign the remaining adjacent channels to the first request in $F$
    else
        Set $k^*$ to the index of the $x$th value of $F$
        if $R_{k^*}/R_i^{\min} < 1$ then
            Find $N_{\mathrm{RB}} = \lceil R_i^{\min}/R_{k^*} \rceil$
            Find the channel block $RB_{c^*}$ that $c^* = \arg\max_c R_{k^*}(RB_c)$
            Assign $N_{\mathrm{RB}}$ consecutive best channels around $RB_c$ on $U_{RB}$ to request $k^*$
        else
            Assign the channel such that $\max_c\{R_{k^*}(RB_c) : RB_c \in U_{RB}\}$ to request $k^*$
        end if
    end if
    $U_{RB} \leftarrow U_{RB} \setminus \{RB's$ assigned in previous step$\}$
    $x \leftarrow x + 1$
end while
```

Algorithm 4 describes in detail the MC-SA allocation procedure. Its initialization phase is identical to the one in the SC-SA algorithm. In addition, it is also identical in the case when the number of requests is smaller than the total number of available RB's.

If the number of requests is higher than the total number of available RB's, the scheduler takes the requests that are experiencing the poorest performance according to metric f_k and try to assign N_{RB} RB's, to do this it find the RB that maximizes the throughput and then look for the adjacent channel (left and right) that has better throughput and repeating this procedure until N_{RB} RB's are assigned to that request.

Finally, once the allocation is performed, the system updates all the relevant parameters.

Note that neither the MC-SA, nor the SC-SA are efficient scheduling algorithms for the case when the number of request that require high throughput is high, as they cannot handle if the total requirements exceeds the capacity of the system, it is up to the admission control strategy to administer the available resources.

4.3 Joint Scheduling and Admission Control

In this section, we propose a new admission control algorithm for LTE Uplink systems that guarantee not only the throughput but also the delay and the traffic priority in the evaluation of the admission of a new request.

Notations

Let K be the set of granted requests, which is partitioned into traffic sets K_i such that $K = \bigcup\limits_{i \in C} K_i$, where C is the index set of the different class of services, and K_i the set of requests of class $i \in C$. Let $R_i^{\min}$ (resp. $d_i^{\max}$) be the minimum bit rate (resp. maximum delay) required by class i of traffic.

We denote by R the average overall throughput per TTI, assuming we maintain statistics over a given number of TTIs, distributed over the time period of concern. Indeed, instantaneous delays and throughputs are computed over a set of n successive TTIs and average values over a set of n_w windows where each window is a set of n successive TTIs.

4.3.1 AC_Fair: A New Admission Control algorithm

To grant a new incoming request, we need to guarantee a minimum rate ($R_i^{\min}$) and a maximum delay ($d_i^{\max}$) for each class i of traffic in addition to the blocking/dropping probabilities. Bit rates and delays are related as, if we increase the average bit rate, denoted by $\overline{R}_i$, by assigning more blocks to some requests, then the average delay, denoted by $\overline{d}_i$, also increases as we need to postpone the block assignment of the remaining requests.

The proposed AC_Fair algorithm checks if the current resource allocation can handle a new request while still satisfying the bit rate and delay requirements of all the active requests and of the new incoming request. Hence, the admission criterion for the new request is expressed as follows:

$$\sum_{i \in C} \frac{\overline{d}_i}{d_i^{\max}} \, |K_i| \, R_i^{\min} + R_{i_{\text{new}}}^{\min} \leq R \tag{26}$$

where $R_{i_{\text{new}}}^{\min} = \left(\overline{d}_i / d_i^{\max}\right) \, R_{i_{\text{new}}}^{\min}$ is the minimum throughput the system needs to provide to the new incoming request taking into account its class of service. It is dependent of the maximum delay that is guaranteed by the system, taking into account the set of already granted requests and the available bandwidth (i.e., available number of blocks or available bandwidth within each block per request). It is assumed that

$$\overline{d}_i \leq d_i^{\max} \qquad \text{and} \qquad \overline{R}_i \geq R_i^{\min}, \tag{27}$$

as otherwise the new incoming request is denied.

We assume that each class i and type of request (new or handover) has a priority Q_j, $j = 1, 2, ..., p^{\max}$ where index 1 corresponds to the highest priority and $p^{\max}$ to the lowest one. We next define the ratio $P_j = P_{\text{bd}_j} / P_{\text{BD}_j}$ where P_{BD} is the target blocking/dropping probability, P_{bd} is the blocking ratio if it is a new request or the dropping ratio if it is a handover request measured in the time frame (of, e.g., n_w windows of TTIs), i.e.,

$$P_{\text{bd}} = \frac{\text{number of rejected requests}}{\text{overall number of requests}} \tag{28}$$

The AC_Fair algorithm can be described as follows:

Parameter $Z \in [0, 1]$ aims to prevent unnecessary blocking request due to previous congestion, for example if we define $Z = 95\%$, and the system load is below that point, it will try to admit

Algorithm 5 AC_Fair

```
Capture physical layer information and calculate average throughput and average delay of class i in the last n msec.
Set $N_i$ (number of new granted requests of class $i$) to zero
Set $Rp$ to $\sum_{i \in C} (\bar{d}_i / d_i^{\max}) \, |K_i| \, R_i^{\min}$
Set $Z$ (priority weight) to a value in $[0, 1]$
Rank incoming requests according to $Q_j$ and select the requests in order of decreasing priority
for each incoming request do
    Collect traffic profile: $P_j, Q_j, R_i^{\min}, d_i^{\max}, \forall i, \forall j$
    if $\bar{d}_i \geq d_i^{\max}$ or $\bar{R}_i \geq R_i^{\min}$ then
        deny the request
    end if
    Find the smallest value of $j*$ such that $P_j* > 0.9$. If none, set $j*$ to $p^{\max}$
    if $Rp < Z \times R$ then
        $j* = p^{\max}$
    end if
    if new incoming request $k$ satisfies condition (26) and if it has priority $Q_j \leq Q_j*$ or priority $P_j > P_{j-1} > P_j*$ then
        Grant request $k$ ; $N_i \leftarrow N_i + 1$
    else
        Reject request $k$
    end if
end for
```

new request without checking the priorities. We assume that if the system is in a relaxed state all requests should have an equal chance of being admitted.

4.3.2 Proposed Scheduling algorithm

The packet scheduling is performed with a one step algorithm, called RA algorithm, that combines time-domain (TD) scheduling and frequency-domain (FD) scheduling. The RA algorithm selects the next request which will be multiplexed in time and frequency with the following metric. For a given request k of class i, it is defined as follows:

$$f_k = \frac{R_k}{\bar{d}_k} \times \frac{d_i^{\max}}{R_i^{\min}} \tag{29}$$

where R_k is the average throughput of request k and $\bar{d}_k$ is the average delay of request k. Indeed, the proposed RA algorithm allocates resource blocks with the aim of maximizing the throughput while making sure that requests will never experience a delay greater than the maximum allowed delay or a throughput smaller than the minimum throughput considering many classes of service.

The reason for using the proportional fair scheduling metric f_k instead of the typical proportional

Algorithm 6 RA Algorithm

```
$U_{RB} \leftarrow$ set of unassigned resource blocks
Define $F = \{f_k \times \delta_k^{\mathrm{d}} \times \delta_k^{R} : k \in K\}$
Sort the values of $F$ in ascending order.
for all $k \in K$ do
    $\delta_k^{\mathrm{d}} = 0$ if $\mathrm{d}_k \geq \mathrm{d}_i^{\max}$ and $k \in K_i$, 1 otherwise
    $\delta_k^{R} = 0$ if $R_k \leq R_i^{\min}$, and $k \in K_i$, 1 otherwise
end for
$x \leftarrow 1$
while $U_{RB} \neq \emptyset$ do
    Set $k^\star$ to the index of the $x$th value of $F$
    Assign $RB_x$ to request $k^\star$
    Assign the channel such that $\max_c\{R_{k^\star}(\mathrm{RB}_c) : \mathrm{RB}_c \in U_{RB}\}$ to request $k^\star$
    $U_{RB} \leftarrow U_{RB} \setminus \{\mathrm{RB}n\}$ ; $x \leftarrow x + 1$
end while
```

fair metric is that this new metric includes the delay requirement by giving more priority to requests that are closer to the maximum delay or to the minimum throughput.

Assuming that the number of requests is always bigger than the number of RBs (as otherwise the block assignment is easy), the RA algorithm assigns RBs based on metric f_k, using the fraction of the current throughput over the minimum throughput and of the maximum delay over the current delay. By doing so, we assign resource blocks to the requests with the most critical delay and throughput.

Chapter 5

Simulation results and Analysis

As described in Chapter 4, we have developed algorithms for single-class scheduling, multi-class scheduling, and a combined solution for admission control and scheduling. Those algorithms are being used in this chapter to investigate their performance.

The chapter is organized as follows. In Section 5.1 we present the simulation model applied in our experiments. It is important to mention that all our simulations were built using MATLAB. In Sections 5.2, 5.3 and 5.4 we present the simulation results and analysis for single-class scheduling, multi-class scheduling and joint scheduling and admission control respectively.

5.1 Simulation model

The simulation model consists of a single cell equipped with an omnidirectional antenna using SC-FDMA uplink based on 3GPP LTE system model. The throughput is averaged over 1000 TTI's, the duration of a TTI is 1msec. The total bandwidth considered is B = 5MHz, subdivided into 25 RBs of 12 sub carriers each with a target BER of $P_b = 10^{-6}$. All mobile users are assumed to transmit at the maximum power considered to be 125mW; the power is equally subdivided among all sub carriers allocated to the mobile in each TTI.

For the throughput calculations, we consider that the data rate is upper bound by the Shannon's formula

$$R_k = \sum_{s \in S_k} \alpha_{s,k} \frac{B}{|S_k|} \log_2[1 + \beta\, \gamma_{s,k}] \tag{30}$$

where β is the SNR gap. For a practical M-QAM system and the theoretical limit based on Shannon capacity [QC99]. Notice that the throughput as defined by the Shannon's formula is expressed at subcarrier level (12 subcarriers = 1 RB), β is given by:

$$\beta = \frac{-1.5}{\ln(5P_b)}. \tag{31}$$

The SNR over a single subcarrier, $\gamma_{s,k}$, is given by:

$$\gamma_{s,k} = \frac{P_{s,k} H_{s,k}}{\sigma_s^2} \tag{32}$$

where $H_{s,k}$ is the channel gain over subcarrier s allocated to user k, σ_s^2 is the noise power, and $P_{s,k}$ is the transmit power, assumed to be subdivided equally among all subcarriers assigned to a user k. Hence

$$H_{s,k}[\text{dB}] = -\kappa - \lambda \log_{10} D_k - \xi_{s,k} + 10 \log_{10} F_{s,k} \tag{33}$$

where κ is the propagation loss chosen to be 128.1 dB, D_k is the distance from the mobile user k to the base station, λ is the path loss exponent set to 37.6, ξ is the log normal shadowing with an 8dB standard deviation, and F is the Rayleigh fading, with a Rayleigh parameter σ such that $E[\sigma^2] = 1$.

All the parameter definitions in (30), (32) and (33) have been expressed assuming subcarrier granularity, and can be easily deduced for resource block granularity, e.g.,:

$$\gamma_{n,k} = \sum_{s \in S_k} \frac{P_{s,k} H_{s,k}}{\sigma_s^2} \tag{34}$$

with S_k as the set of subcarriers allocated to user k.

$$S_k = \{s \in S : s \in N\}, \tag{35}$$

We also use the Jain's fairness index [JCH84]:

$$\text{FI}_{\text{Jain}} = \frac{\left[\sum_{k \in K} R_k\right]^2}{K \times \sum_{k \in K} R_k^2} \tag{36}$$

Parameter	Setting
System Bandwidth	5MHz
Subcarriers per resource block	12
Number of Resource Blocks	25
Minimum throughput	200kbps
Maximum delay	10ms, 20ms, 40ms
Traffic model	Infinitely backlogged
Transmission Time Interval	1ms
Number of users in cell	8, 24, 48, 96, 144
Noise power per Hz	160dBm
User distribution	uniform
Channel model	Urban
Type of system	Single cell

Table 2: Simulation parameters for single-class scheduling

where $\mathrm{FI}_{\mathrm{Jain}}$ is the actual data rate achieved by user k with K users in the system.

Single-class scheduling

Table 2 summarizes a list of the default simulation parameters and assumptions used for the Single-class scheduling experiments.

Multi-class Scheduling

Table 3 summarizes a list of the default simulation parameters and assumptions used in the Multi-class scheduling experiments. To simplify the analysis without loss of generality, we have considered three service classes (e.g., voice, video and web). The traffic model settings are given in Table 4, and the traffic distribution is shown in Table 5. Requirements of packet delays are considered as one fifth of the end to end packet delays in Table 4. We also considered a very demanding class to oblige the system to assign more than one RB to a request, i.e., the throughput requirement of video for standard definition television.

Joint Scheduling and Admission control

The main simulation parameters are listed in Table 6. All the requests are generated in the system according to a Poisson arrival process. If the AC decision criterion proposed in Section 4.3 is fulfilled, the request is admitted, otherwise the request is blocked.

Parameter	Setting
System Bandwidth	5MHz
Subcarriers per resource block	12
Number of Resource Blocks	25
Traffic model	Infinitely backlogged
Transmission Time Interval	1ms
Number of users in cell	10, 20, 30, 40, 50, 60, 70
Noise power per Hz	160dBm
Channel model	Urban
Type of system	Single cell
User maximum power	125 mW
Available MCSs	M-QAM
Simulation time	10000 TTIs

Table 3: Simulation parameters for multi-class scheduling

Services	Delay budget	Rate budget	Packet loss rate
Voice	100 ms	64Kbps	10^{-2}
Video	150 ms	3Mbps	10^{-3}
Web	300 ms	128Kbps	10^{-6}

Table 4: QoS requirements for multi-class scheduling

Services	Distribution
Voice	50%
Video	10%
Web	40%

Table 5: Traffic distribution for multi-class scheduling

Parameter	Assumption
System bandwidth	5 MHz (25 PRBs, 180 kHz per PRB)
Path loss	128.1 + 37.6 log10(d in km) dB
Log-normal shadowing	Standard deviation = 8 dB
Shadowing correlation	1.0 for intra-site, 0.5 for inter-site
Penetration loss	20 dB
Fast fading	Typical Urban (TU3)
TTI	1 ms
User maximum power	125 mW
Available MCSs	M-QAM
Request arrival	Poisson process
Request arrival rates	[0.9, 0.96, 1.02, 1.08, 1.14, 1.2] calls/s
Average call duration	300 s
Simulation time	3600 s
Z Parameter	95 %

Table 6: Parameters

Services	QCI	Priority	Delay budget	Rate budget
Voice	1	2	100 ms	64Kbps
Video	2	4	150 ms	256Kbps
Web	7	6	300 ms	16Kbps

Table 7: QoS requirements

Services	Distribution
Handover Voice	8%
Handover Video	4%
Handover Web	8%
Voice	32%
Video	16%
Web	32%

Table 8: Traffic distribution

To simplify the analysis without loss of generality, we consider three service classes (voice, video and web). Furthermore, the calls from the same service class can be either new in the cell or a handover from another cell, with handover calls having higher priority than new incoming calls as it is more annoying to drop a call than not having access to the network due to congestion; it might be also closer to the best conditions of the service level agreement (SLA). We choose arbitrarily the following priority scheme for our simulations handover voice > handover video > handover web > new voice > new video > new web, with handover voice as the highest priority and new web as the lowest priority. The traffic model settings are given in Table 7

Requirements of packet delays are considered as one fifth of the end to end packet delays in Table 1. In simulations, AC functionality is tightly coupled with scheduling algorithm as the QoS management is made in the scheduling part and priorities are handle by the AC algorithm. We consider three service classes, the traffic distribution is show in Table 8.

Since many admission control algorithms have been proposed for cellular systems, it is difficult to directly compare the performance of existing AC algorithms and the proposed AC algorithm. Here, we compare performance of our AC algorithm with that of the reference AC [MA08], which can be expressed as follows:

$$\sum_{k=1}^{K} R_k^{\min} + R_{k^{new}}^{\min} \leq R^{max} \tag{37}$$

where $R^{\min}$ is the minimum throughput required and R^{max} is a predefined throughput that can be manually tuned to maximize efficiency the reference AC decides to admit a new request if the sum of the $R^{\min}$ of the new and the existing uses is less than or equal to R^{max}

5.2 Single-class Scheduling

Results reported in this section have been obtained using a logarithmic throughput utility ($\ln(R)$, see expression (14)) in the Low Complexity Algorithm (LC) and the LC-Delay algorithm. Proportional Fair (PF), and Proportional Fair with Delay constraint (PF-Delay) uses an equivalent form of the logarithmic approach r_i/R, see expression (16).

All simulations have been conducted with the parameters described in Section 4.1. We analyze the performance of the algorithms in terms of allocation vs. distance, throughput, delay, and fairness to assess the quality of their scheduling and resource block allocation.

Users served

In Figure 8, we compare the average number of effectively served users vs. the total number of users in the system. Note that there are users in the system that are not assigned resources or such that their delay and throughput constraints are not met. As the number of users increases, most of the algorithms reach their maximum capacity and then their performance starts to deteriorate. We observe that the PF-Delay algorithm clearly outperforms the others, with a performance peak at around 144 users (0% blocking).

The reason for the performance reduction, after the maximum has been reached, showed in the figure is that all the proposed algorithms in their effort to serve new users, take some resources from the existing users, driving some of them not to fulfill the minimum requirements. This fact justifies the need for admission control schemes that allows the system to reach their maximum capacity.

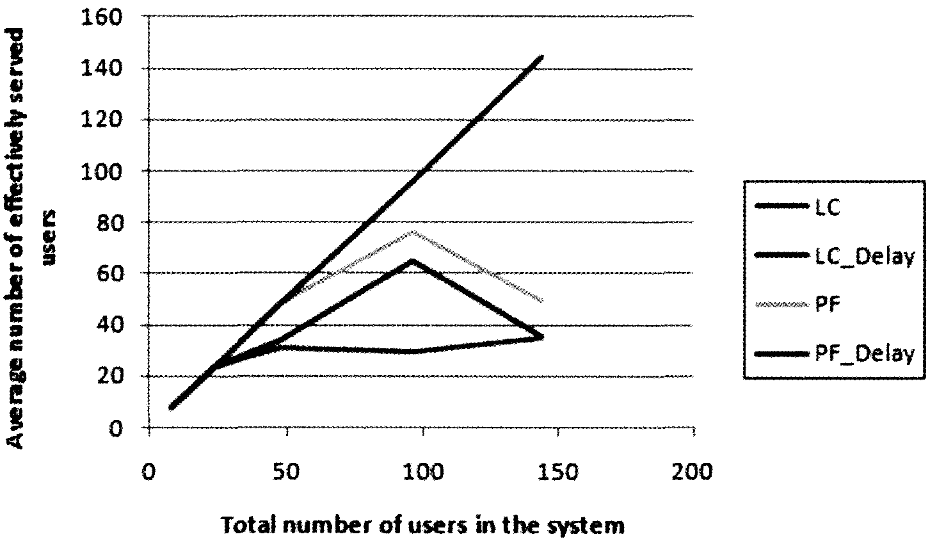

Figure 8: Total number of users in the system vs. Average number of users effectively served

Delay Requirement

One of the main criteria of real-time requests is the delay requirement. We conducted experiments in the context of voice requests for which, according to [3GP07a] (see also [JWMT07]), the radio interface delay budget is 50ms (note that, in the worst case, the call may originate and terminate in an LTE cell). Indeed, see [JWMT07], a VoIP user is in outage (unsatisfied) if more than 2% of its packets in a 60s call cannot be correctly received within 50ms. The system capacity is defined as the maximum load which more than 95% of the users are satisfied.

In Figure 9, we show the variation of the delay as the number of users increases (maximum tolerated delay has been set to 20ms). Clearly PF-Delay outperforms the others. Indeed, all three other algorithms perform similarly, with a gap that is significantly increasing as the number of users is growing, with respect to the average number of satisfactorily served users.

Fairness

To obtain an indication about the user fairness of scheduling provided by the different algorithms, we analyze the short-term fairness. We use a fairness index measured by the data rate fairness criterion (Fairness = 1 means that all users receive equal data rate within time t) proposed by [JCH84], see

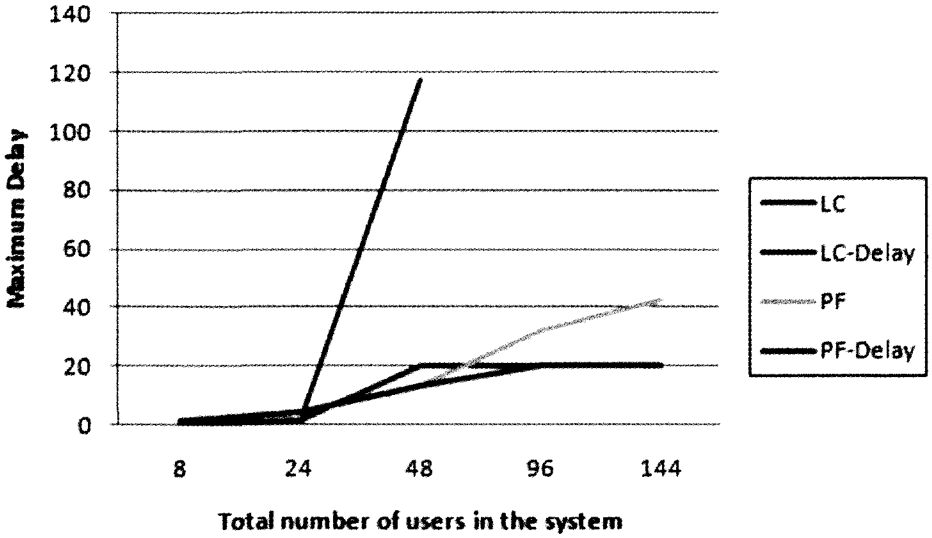

Figure 9: Maximum delay vs. Number of users

Equation 36 defined in section 5.1.

We analyze the fairness by varying the number of users. In Figure 10, we depict the value of the Fairness index as a function of the number of users. While the values of the fairness index of the four algorithms are similar for a small number of users, we observe that, as the number of users increases, PF-Delay remains as fair as PF, while satisfying the delay constraint.

Resource Allocation

We first measure the average subcarrier allocation vs. distance for the four algorithms as shown in Figure 11. For a small number of users ($k < RB$), all algorithms have similar results. For a large number of users, we can see that LC algorithm assigns few resources to users in the edge, resulting a priori with the possibility of greater delay. Algorithm LC-Delay solves this issue by constraining the maximum delay and assigning the best channel. Algorithm PF and PF-Delay keeps performing the assignment proportional even with a large number of users.

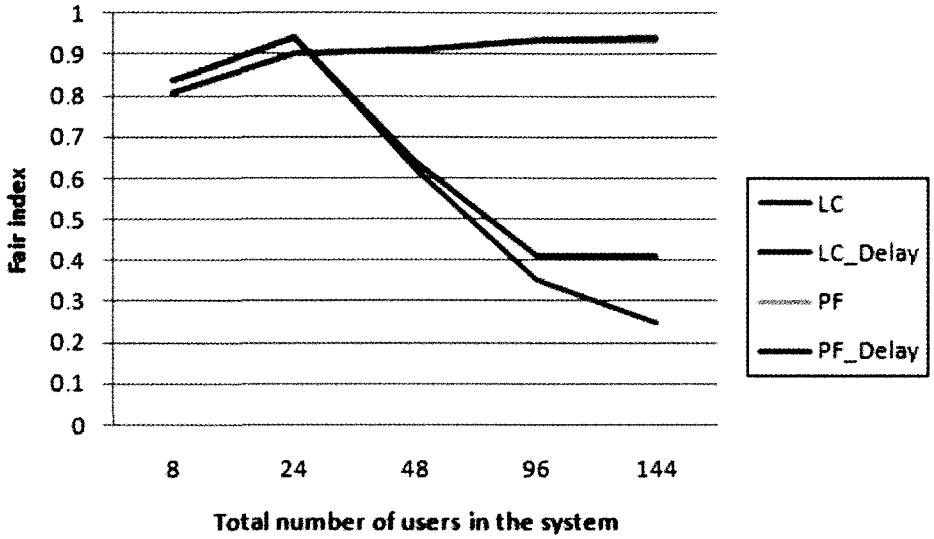

Figure 10: Fairness vs. Number of users

Throughput

The user throughput for 64 users is shown in Figure 12 (maximum delay = 20ms) and numerical results for various number of users and delays are shown in Table 9. It can be seen from Table 9 that PF-Delay algorithm outperforms the others. A remarkable observation is that when the number of users is small, all algorithms performs almost equally but as the number of users increase, our proposed algorithm achieves a better performance, allowing to have around 3 times the maximum number of served users than the other algorithms, at the expense of a little increase in complexity.

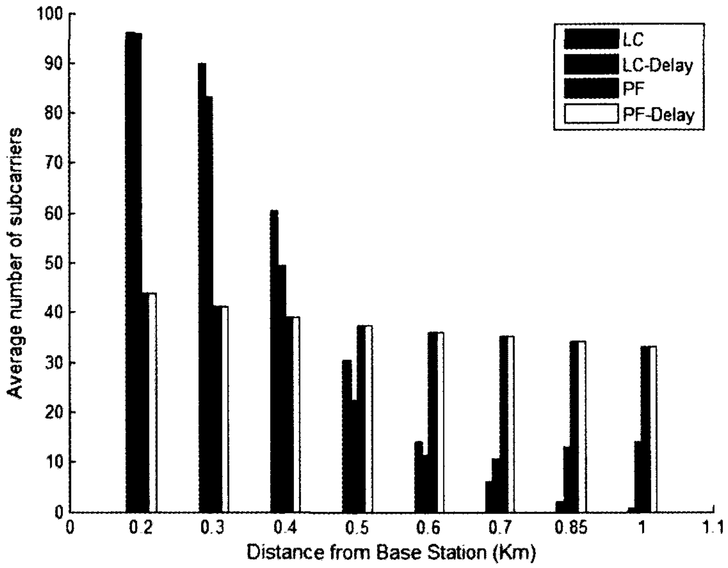

Figure 11: Average subcarrier allocation vs. distance

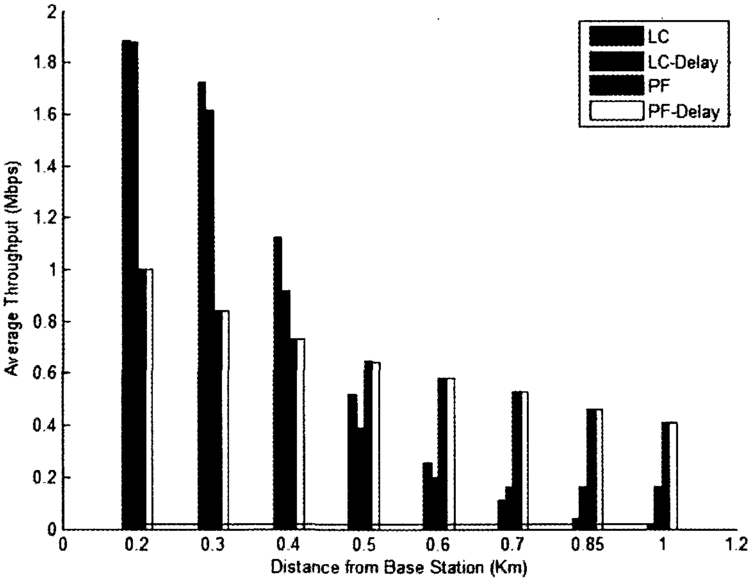

Figure 12: Average throughput achieved by each user vs. distance

		Maximum Delay		
Algorithm	# users	10	20	40
LC	8	28.0	27.9	27.9
LC_Delay		28.0	27.9	27.9
PF		28.1	28.2	28.2
PF_Delay		28.1	28.2	28.2
LC	24	36.6	36.6	36.7
LC_Delay		36.6	36.6	36.7
PF		43.7	43.7	43.6
PF_Delay		43.7	43.7	43.6
LC	48	*	*	*
LC_Delay		51.0	51.4	51.6
PF		*	50.2	50.2
PF_Delay		50.2	50.2	50.2
LC	96	*	*	*
LC_Delay		*	55.5	55.5
PF		*	*	53.9
PF_Delay		53.1	54.0	53.9
LC	144	*	*	*
LC_Delay		*	*	*
PF		*	*	*
PF_Delay		49.5	55.6	55.6

Table 9: Average sum throughput (Mbps)

5.3 Multi-class Scheduling

Results reported in this section have been conducted with the parameters described in the previous subsection. We analyze the performance of the algorithms in terms of throughput, and delay to assess the quality of their scheduling and resource block allocation.

In Figure 13, we show the ratio of the average satisfied users versus the total number of users in the system. Note that although we are assigning resources to all users, not all are meeting the QoS requirements.

As we can see from the simulation result, algorithm MC-SA has around 5% better performance than SC-SA, this difference is mainly due to the video requests that are the ones who consumes more bandwidth and needs more than one RB in a given TTI.

Figure 14 shows the average throughput of video calls versus the total number of requests in the system, it can clearly be seen from the figure that MC-SA satisfies the throughput requirement by assigning more than one RB, while SC-SA is not capable to fulfill the video call throughput requirements since it only assigns one RB every TTI.

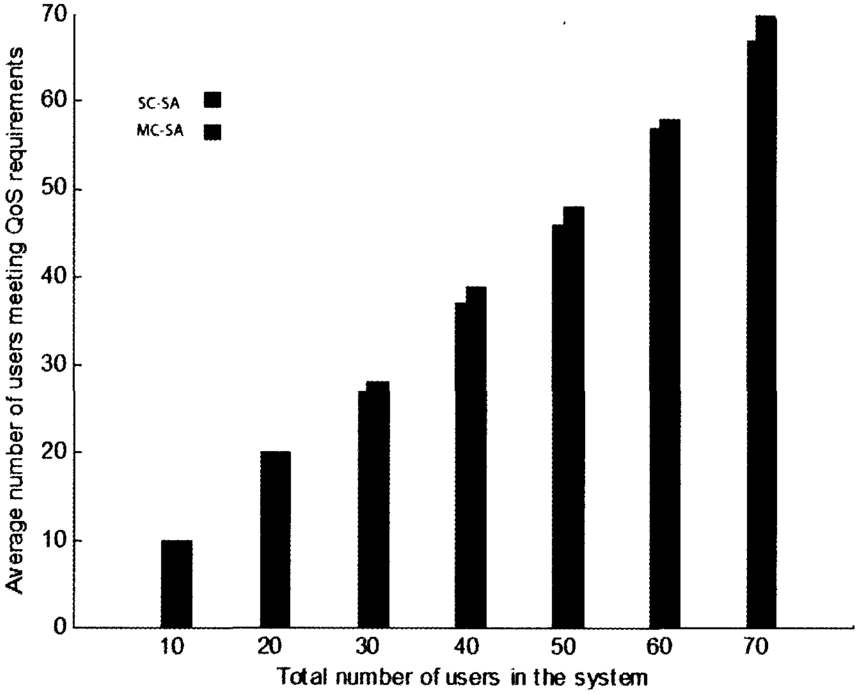

Figure 13: Number of users in the system vs. number of users effectively served

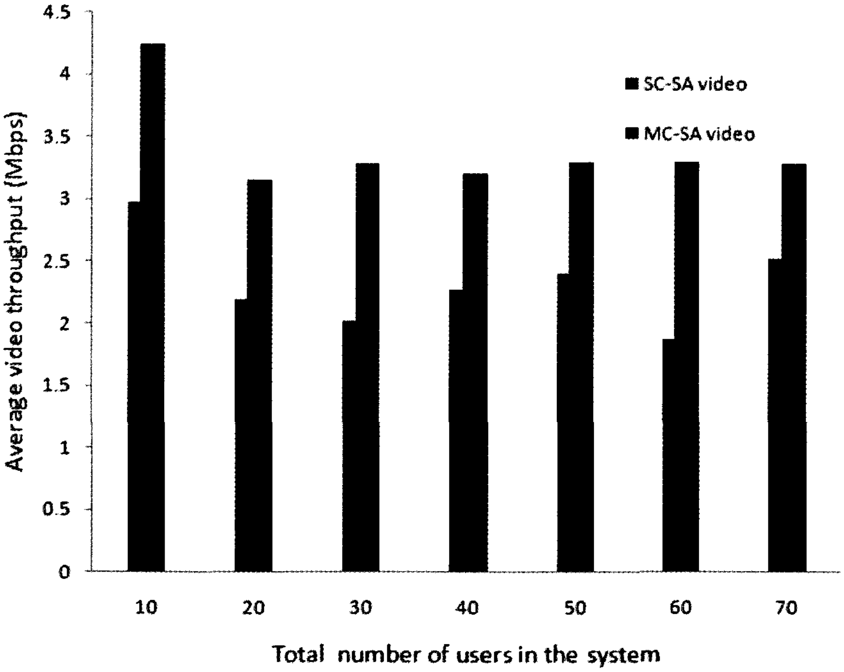

Figure 14: Average video throughput

Figure 15 and Figure 16 show the average voice throughput and average web throughput as a function of the total number of requests in the system. Both algorithms satisfy the throughput requirement, even though MC-SA has slightly lower performance. Indeed, this is because MC-SA uses this throughput to satisfy the video call demand.

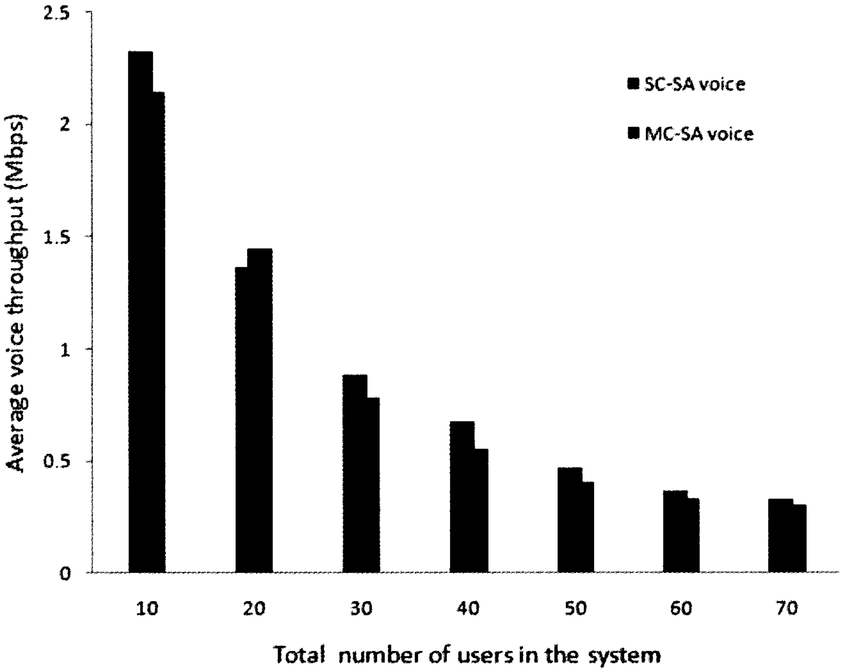

Figure 15: Average voice throughput

Figure 17 shows the average delay of each class versus the total number of requests in the system, it can be seen from the figure that MC-SA and SC-SA meet the maximum allowed delay. One important thing we can point out is that in order to meet the video requirements, both systems assign resources to video calls at every TTI, nevertheless MC-SA is the only one which guarantees the minimum throughput.

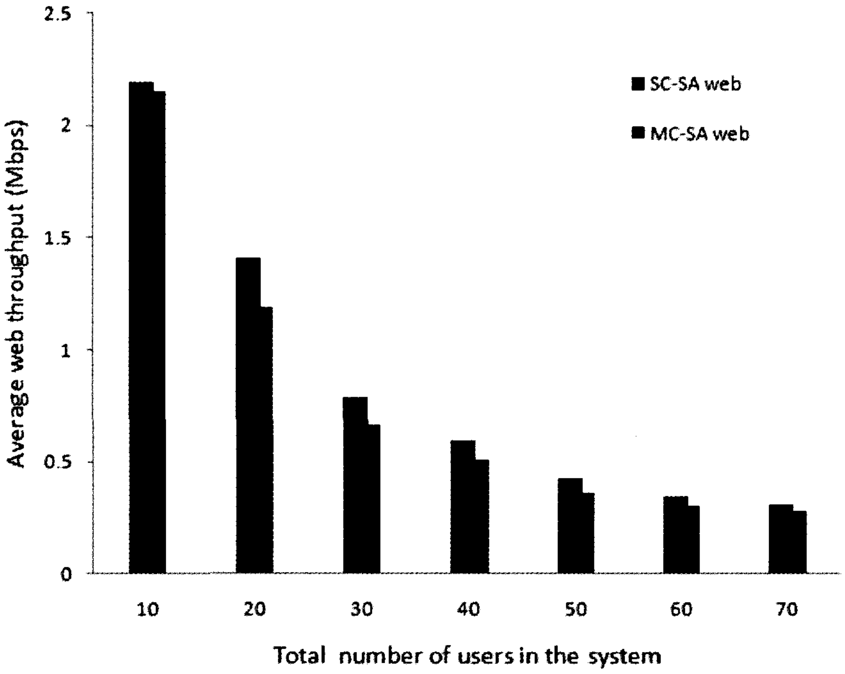

Figure 16: Average web throughput

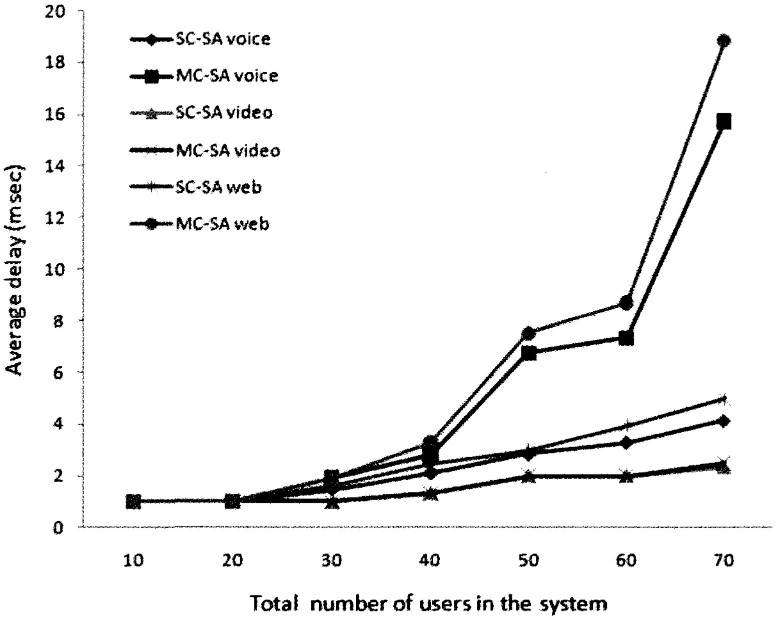

Figure 17: Average voice delay

Figure 18 shows that if we do not consider the delay as part of metric f we can reach the maximum delay sooner.

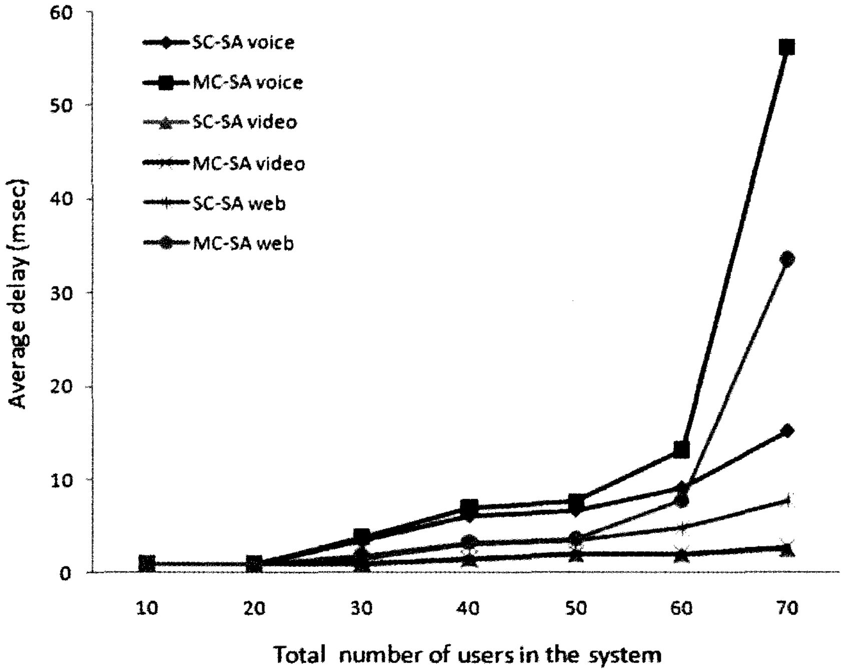

Figure 18: Average voice delay excluding the delay parameter in f

5.4 Joint Scheduling and Admission Control

The performance of the proposed AC and resource allocation algorithms is evaluated using blocking and outage probabilities as well as the average number of accepted requests in the system and the average sum throughput in the cell. Blocking probability (P_b) is defined as the ratio of the number of blocked requests to the total number of requests. Outage probability (P_o) is calculated as the ratio of the number of requests not fulfilling their throughput and delay requirements to the total number of admitted requests.

Figure 19 and Figure 20 show the blocking probabilities for our proposed AC algorithm and the reference AC (see expression (37)) as a function of the call arrival rate. From the figure, we can draw the following conclusions. Firstly, the blocking probability increases as the call arrival rate increases. Secondly, the blocking probability is much higher for video calls as this class of traffic is more demanding in terms of bandwidth. Finally and most important in Figure 19, we notice that blocking probabilities follow the required priority scheme (ex. gold > silver > bronze), while, in Figure 20, the priority distribution is random. This is because the blocking probability is not only dependent of the maximum capacity of the system but also of the throughput that the new requests are asking to the system.

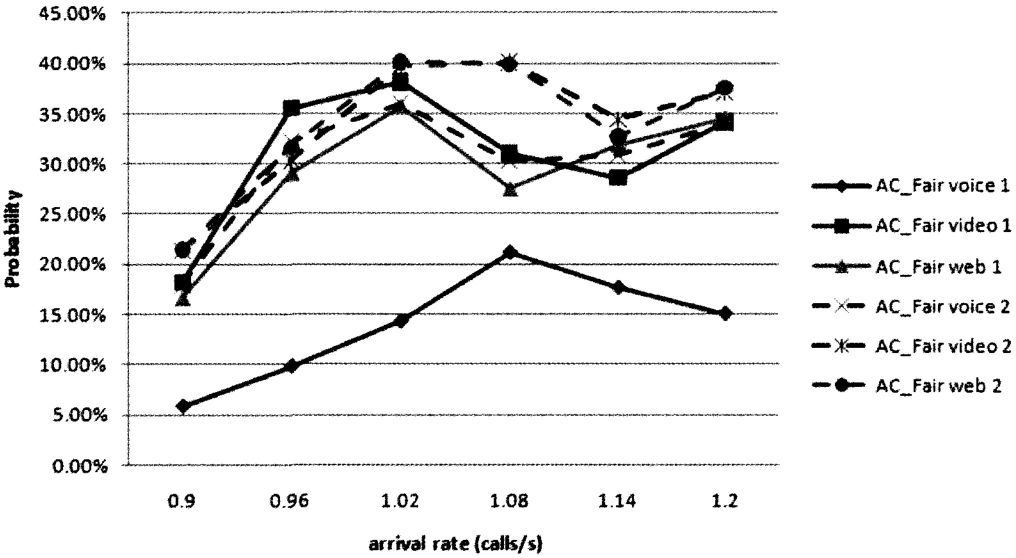

Figure 19: AC_Fair algorithm blocking probabilities

Figure 21 depicts the outage probabilities versus the call arrival rate, it can be seen from the figure that the outage probabilities is best for the AC_Fair algorithm than for the reference AC. This is due to the fact that AC_Fair admits a new user only if its QoS can be satisfied, we must remember that in order to evaluate the performance of any admission control algorithm, we have to generate enough traffic to lead the system to the congestion region.

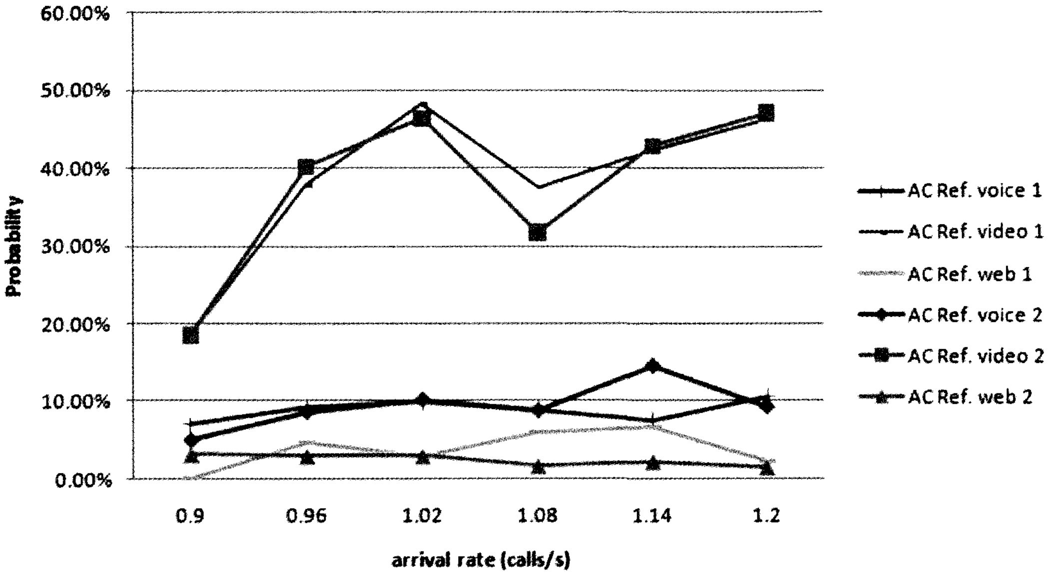

Figure 20: Reference algorithm blocking probabilities

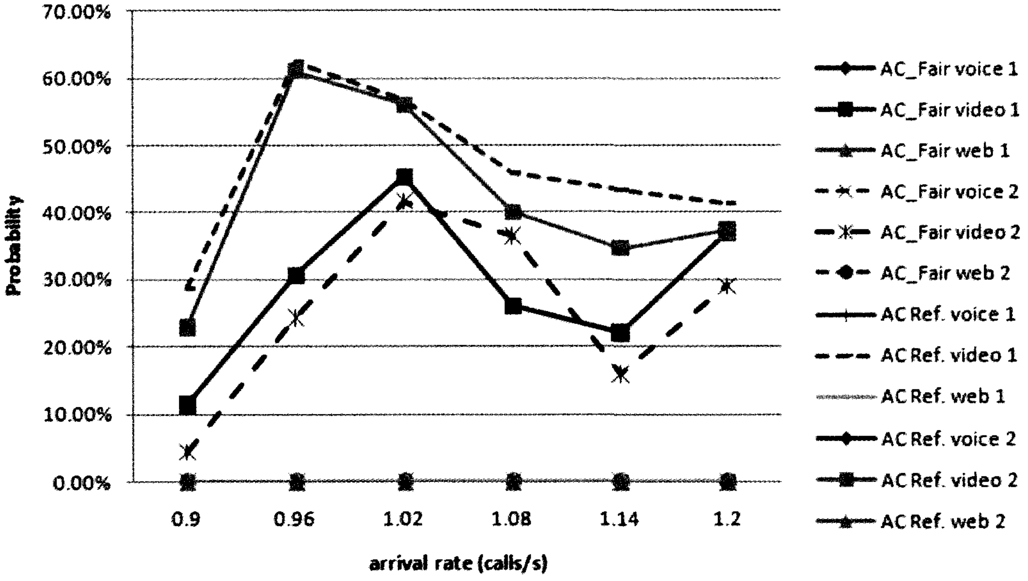

Figure 21: Outage probability

Figure 22 shows the average number of requests admitted in the system versus the arrival rate, we also consider the average number of request fulfilling the QoS requirements (useful number of request), for AC_Fair the values of the average number of requests fulfilling the QoS requirements are slightly below of that of the reference AC algorithm, this is due to the fact that in the priority evaluation it can be loss some efficiency.

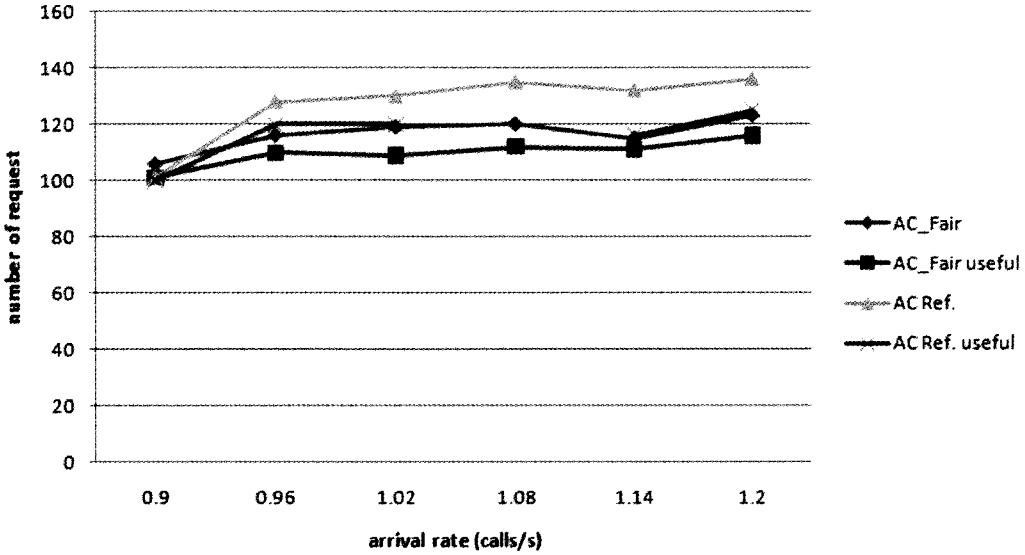

Figure 22: Average number of admitted requests

Figure 23 shows the useful sum throughput in the cell, meaning the sum throughput of all the request fulfilling the QoS requirements, it can be seen from the figure that AC_Fair algorithm has about the same behavior than the reference AC, values varies on a maximum of 2%. It is important to note that the maximum Rate *Rmax* of the reference AC algorithm is a manually tunable parameter and we are choosing the best value to make fair comparisons with our proposed AC algorithm. On the contrary AC_Fair adapts dynamically to the traffic conditions.

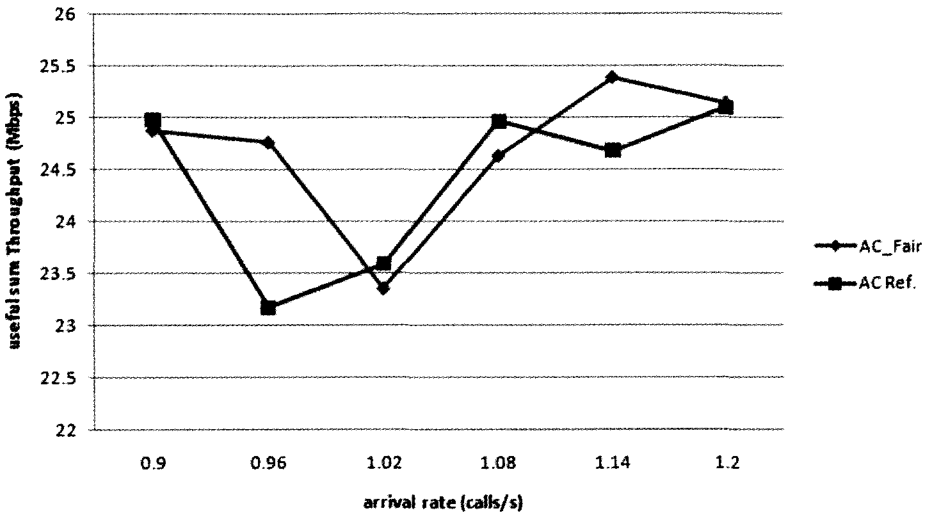

Figure 23: Average useful sum throughput

Figure 24 compares the useful sum throughput of video calls versus the arrival rate, it can be seen from the figure that the useful throughput of video calls for AC_Fair algorithm is higher than the reference AC algorithm. This is because the outage probability of video calls is significantly higher, mostly due to the fact that, in our experiments, video calls are more demanding (higher data rate).

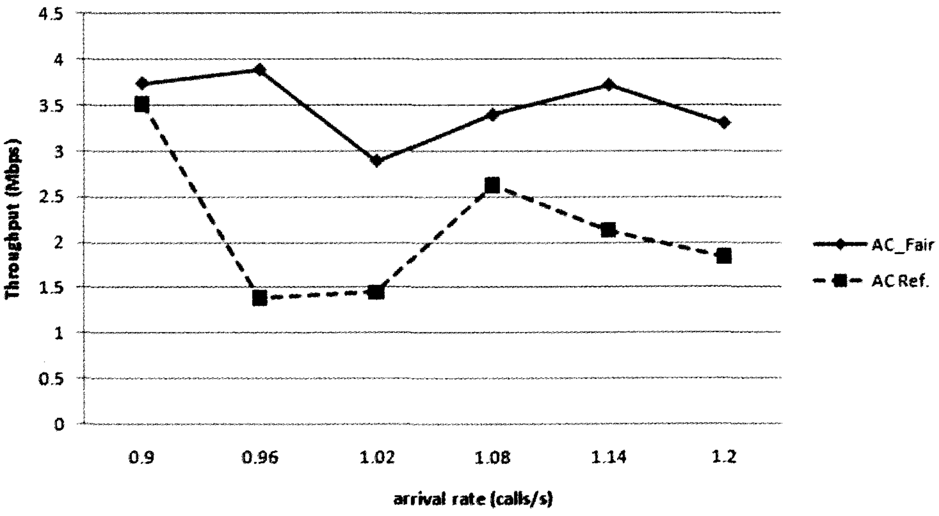

Figure 24: Average video useful sum throughput

5.5 Conclusions of simulations

Single-class Scheduling

Uplink scheduling for single cell LTE systems was analyzed, expressing for the first time the adjacent resource block constraints, together with the maximum delay and minimum throughput constraints. Due to the difficulty of solving efficiently the mathematical model, we turn our attention to greedy heuristics and proposed two highly scalable heuristics. Their performances were analyzed not only in terms of throughput, resource allocation and fairness but also in terms of delay and number of users effectively served. We also compare our results with two very recently proposed heuristics. Results show that our new heuristics performed much better than the previously proposed heuristics, both in terms of the number of served users, but also with respect to the delay constraints and the user fairness.

Multi-class Scheduling

We developed two Resource Allocation and Scheduling algorithms for handling multiclass QoS in LTE Uplink systems. The Resource Allocation assigns RB's in a fair way such that the throughput and delay are adaptively adjusted according to the traffic load. System performance was evaluated using simulations. Results show that although assigning more than one resource block as in MC-SA helps to serve requests with higher throughput requirements, it also has a negative effect by reducing the throughput of the other services (voice, web); as compared with a SC-SA system that only assigns one resource block per TTI to a given request. Nevertheless MC-SA outperforms SC-SA by enabling to operate with higher demands of throughput, and maintaining the basic QoS requirements for all admitted requests.

Joint Scheduling and Admission control

We developed a combined Admission Control and Resource Allocation algorithm for handling multiclass GoS and QoS in LTE Uplink systems. The AC determines if a new request can be accepted based on its priority and in the real minimum throughput and delay the system can offer so as to fulfill the QoS of the new and existing requests, and the Resource Allocation assigns RB's in a fair way such that the throughput and delay is adaptively adjusted according to the traffic load. System performance is evaluated using simulations. Results show that although the total sum throughput is not improved, the proposed AC algorithm gain for the most sensitive traffic can be around 20% over the reference AC algorithm without sacrificing the overall system capacity and at the same time guarantying GoS and maintaining the basic QoS requirements for all admitted requests.

Chapter 6

Conclusion

6.1 Summary of the Thesis

An LTE uplink system, as any other mobile technology, has some requirements, i.e., improves the end-to-end QoS. To achieve these objectives, LTE relies not only on new technologies like SC-FDMA, but also on improved resource management. By analyzing and exploiting the specific characteristics of the admission control and scheduling, we can find additional opportunities to improve the user's experienced Grade of Service and Quality of Service.

In this thesis, first we have investigated scheduling techniques for single-class and multi-class QoS. Regarding scheduling, we have formulated a new optimization model for single-class and multi-class environments that includes the LTE specific constraints, i.e., contiguity and delay constraints; we have also developed four heuristics that include the delay as an important parameter. All these algorithms were analyzed in terms of throughput, delay and fairness,

We also have investigated admission control for multi-class systems with a priority scheme. In this thesis we have developed an algorithm that achieves QoS and GoS by taking advantage of the scheduling design. Furthermore, this new design is capable of adaptively adjust to different traffic loads. This research relies on simulations.

In summary, the benefits of multi-class admission control and scheduling over single cell systems are increased system capacity and overall provisioned quality of service. Furthermore, it has been shown that the delay plays an important role when developing scheduling algorithms. Such a principle has been adapted to rate control in the admission control mechanisms, and it allows us to have a more refined multi-class admission control, leading to improve the system capacity and QoS.

6.2 Future Work

In Sections 4.1 and 4.2, we developed an optimization model for single-class and multi-class scheduling. Although we properly describe the optimization model, we did not solve it. It is worthwhile to find the optimal solution of these models to have a real measurement of how close our heuristic solutions are to the optimal values.

All the heuristics that we developed in this thesis for the scheduling approach were for single-cell systems. It is necessary to investigate the behavior of our heuristics for multi-cell systems.

It would be also interesting to develop an optimization model for multi-time slot that allows us to assign resources for more than one TTI. Although it will not be practical for a real time solution, it could allow as having a planning tool.

In the admission control part, it will be interesting to have a multi-cell model that allows us to evaluate if we can take advantage of the information of other cells to increase the system capacity and possibly reduce the blocking probabilities as users may have service from a less congested cell.

In this thesis, we had focused on uplink LTE systems. Nevertheless, we would like to investigate on a combined admission control strategy for downlink and uplink. As in IP based networks the traffic is mostly asymmetric, we need to decide whether we can accept or reject a new request based on the information of both directions.

Bibliography

[3GP07a] 3GPP. *LTE Physical Layer Framework for Performance Verification*, February 2007. 3GPP TSG-RAN WG1 Meeting #48 R1-070674.

[3GP07b] 3GPP. *Way Forward on Power Control of PUSCH*, June 2007. 3GPP TSG-RAN WG1 #49-bis.

[3GP09] 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification*, December 2009. 3GPP TS 36.323 v9.0.0.

[3GP10a] 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN);Overall description*, March 2010. 3GPP TS 36.300 v8.12.0.

[3GP10b] 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); LTE physical layer; General description*, March 2010. 3GPP TS 36.201 v9.1.0.

[3GP10c] 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification*, March 2010. 3GPP TS 36.321 v9.2.0.

[3GP10d] 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) protocol specification*, March 2010. 3GPP TS 36.322 v9.1.0.

[3GP10e] 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification*, March 2010. 3GPP TS 36.331 v9.2.0.

[3GP10f] 3GPP. *Policy and charging control architecture*, March 2010. 3GPP TS 23.203 v9.4.0.

[AQS05] M. Andrews, L. Qian, and A. Stolyar. Optimal utility based multi-user throughput allocation subject to throughput constraints. In *IEEE Annual Joint Conference of the IEEE Computer and Communications Societies - INFOCOM*, volume 4, pages 2415–2424, March 2005.

[ARC+08] M. Anas, C. Rosa, F.D. Calabrese, P.H. Michaelsen, K.I. Pedersen, and P.E. Mogensen. QoS-Aware Single Cell Admission Control for UTRAN LTE Uplink. In *IEEE Vehicular Technology Conference (VTC)*, pages 2487 –2491, May 2008.

[BCC+09] S. Jae Bae, B. Choi, M. Young Chung, J. Ju Lee, and S. Kwon. Delay-aware call admission control algorithm in 3GPP LTE system. *TENCON, IEEE Region 10 Conference*, pages 1–6, November 2009.

[CG07] S.B. Chaudhry and R.K. Guha. Adaptive connection admission control and packet scheduling for QoS provisioning in mobile WiMAX. In *IEEE International Conference on, Signal Processing and Communications. (ICSPC)*, pages 1355 –1358, 2007.

[DJ10a] O. Delgado and B. Jaumard. Joint Admission Control and Resource Allocation with GoS and QoS in LTE Uplink. In *IEEE International Conference on, BWA GLOBECOM*, 2010. Submitted.

[DJ10b] O. Delgado and B. Jaumard. Scheduling and Resource Allocation for Multiclass Services in LTE Uplink systems. In *6th IEEE International Conference on, Wireless and Mobile Computing, Networking and communication (WiMob)*, 2010.

[DJ10c] O. Delgado and B. Jaumard. Scheduling and Resource Allocation in LTE Uplink with a Delay Requirement. In *IEEE Eighth Annual Communication Networks and Services Research Conference (CNSR)*, pages 268 –275, May 2010.

[DPSB08] E. Dahlman, S. Parkvall, J. Skold, and P. Beming. *3G Evolution, Second Edition: HSPA and LTE for Mobile Broadband.* Academic Press, 2008.

[GC08] L. Gao and S. Cui. Efficient Subcarrier, Power, and Rate Allocation with Fairness Consideration for OFDMA Uplink. *IEEE Transactions on Wireless Communications*, 7(5):1507–1511, May 2008.

[HA09] Holma H. and Toskala A. *LTE for UMTS - OFDMA and SC-FDMA Based Radio Access*. Wiley, June 2009.

[Hos03] P. Hosein. A Class-Based Admission Control algorithm for shared wireless channels supporting QoS services. In *Fifth IFIP-TC6 Conference on Mobile and Wireless Communication Networks*, pages 81 – 85, 2003.

[HSAB09] J. Huang, V.G. Subramanian, R. Agrawal, and R. Berry. Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks. *IEEE Journal on Selected Areas in Communications*, 27:226–234, February 2009.

[JCH84] R. Jain, D. M. Chiu, and W. Hawe. A quantitative measure of fairness and discrimination for resource allocation in shared systems. DEC Research Report TR-301, Eastern Research Lab, 1984.

[JS07] L. Jing and S. Sampalli. Cell mobility based admission control for wireless networks with link adaptation. In *IEEE International Conference on Communications (ICC)*, pages 5862 –5867, August 2007.

[JWMT07] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala. Principle and performance of semi-persistent scheduling for VoIP in LTE system. In *International Conference on Wireless Communications, Networking and Mobile Computing (WiCom)*, pages 2861–2864, September 2007.

[KHK05] K. Kim, Y. Han, and S.-L. Kim. Joint subcarrier and power allocation in uplink OFDMA systems. *IEEE Communications Letters*, 9:526 – 528, June 2005.

[KMS08] S. Kalikivayi, I.S. Misra, and K. Saha. Bandwidth and Delay Guaranteed Call Admission Control Scheme for QOS Provisioning in IEEE 802.16e Mobile WiMAX. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1 –6, November 2008.

[KSC09] Y. Kim, K. Son, and S. Chong. QoS Scheduling for Heterogeneous Traffic in OFDMA-Based Wireless Systems. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1 –6, November 2009.

[KW04] H. Kushner and P. Whiting. Convergence of proportional-fair sharing algorithms under general conditions. *IEEE Transactions on Wireless Communications*, 3:1250 – 1259, July 2004.

[LPM+09] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu. Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink. In *IEEE Annual Joint Conference of the IEEE Computer and Communications Societies - INFOCOM*, pages 2611 – 2615, 2009.

[LYZ+08] H. Lei, M. Yu, A. Zhao, Y. Chang, and D. Yang. Adaptive Connection Admission Control Algorithm for LTE Systems. pages 2336 –2340, May 2008.

[MA08] F. D. Calabrese K. I. Pedersen P. E. Mogensen M. Anas, C. Rosa. Combined Admission Control and Scheduling for QoS Differentiation in LTE Uplink. *IEEE 68th Vehicular Technology Conference (VTC)*, pages 1–5, September 2008.

[MLG06] H.G. Myung, J. Lim, and D.J. Goodman. Single carrier FDMA for uplink wireless transmission. *IEEE Vehicular Technology Magazine*, pages 30–38, September 2006.

[MPKM08] G. Mongha, K.I. Pedersen, I.Z. Kovacs, and P.E. Mogensen. QoS Oriented Time and Frequency Domain Packet Schedulers for The UTRAN Long Term Evolution. In *IEEE Vehicular Technology Conference (VTC)*, pages 2532 –2536, May 2008.

[NS08] C. Y. Ng and C. W. Sung. Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems. *IEEE Transactions on Wireless Communications*, 7(5):1667–1675, May 2008.

[PKF+09] K.I. Pedersen, T.E. Kolding, F. Frederiksen, I.Z. Kovacs, D. Laselva, and P.E. Mogensen. An overview of downlink radio resource management for UTRAN long-term evolution. *IEEE Communications Magazine*, 47(7):86 –93, July 2009.

[QC99] X. Qiu and K. Chawla. On the performance of adaptive modulation in cellular systems. In *IEEE Transactions on Communications*, volume 47, pages 884 – 895, 1999.

[QHS+09] M. Qian, Y. Huang, J. Shi, Y. Yuan, L. Tian, and E. Dutkiewicz. A Novel Radio Admission Control Scheme for Multiclass Services in LTE Systems. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1 –6, November 2009.

[QRTC09] Y. Qian, C. Ren, S. Tang, and M. Chen. Multi-service QoS guaranteed based downlink cross-layer resource block allocation algorithm in LTE systems. In *IEEE International Conference on Wireless Communications Signal Processing*, pages 1 –4, 2009.

[RIK10] M.M. Rana, M.S. Islam, and A.Z. Kouzani. Peak to Average Power Ratio Analysis for LTE Systems. pages 516 –520, 2010.

[RQL07] B. Rong, Y. Qian, and K. Lu. Downlink Call Admission Control in Multiservice WiMAX Networks. In *IEEE International Conference on Communications (ICC)*, pages 5082 –5087, 2007.

[SCK10] A. Saddoud, L.F. Chaari, and L. Kamoun. Admission control scheme and bandwidth allocation for mobile WiMAX networks. In *Informatics and Systems (INFOS), 2010 The 7th International Conference on*, pages 1 –5, 2010.

[STB09] S. Sesia, I. Toufik, and M. Baker. *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.

[TC08] C. Tarhini and T. Chahed. Density-based admission control in IEEE802.16e Mobile WiMAX. In *1st IFIP, Wireless Days (WD)*, pages 1 –5, 2008.

[TJP08] A. Teh, A. Jayasuriya, and P. Pudney. Admission control in wireless infrastructure networks based on the predicted percentage of delayed packets. In *14th Asia-Pacific Conference on Communications (APCC)*, pages 1 –5, October 2008.

[TLT+07] K.F. Tsang, L.T. Lee, H.Y. Tung, R. Lam, Y.T. Sun, and K.T. Ko. Admission control scheme for mobile WiMAX networks. In *IEEE International Symposium on Consumer Electronics (ISCE)*, pages 1 –5, 2007.

[TTLK08] H.Y. Tung, K.F. Tsang, L.T. Lee, and K.T. Ko. QoS for Mobile WiMAX Networks: Call Admission Control and Bandwidth Allocation. In *5th IEEE Consumer Communications and Networking Conference (CCNC)*, pages 576 –580, 2008.

[WWS07] M. Wemersson, S. Wanstedt, and P. Synnergren. Effects of QoS Scheduling Strategies on Performance of Mixed Services over LTE. In *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1 –5, 2007.

[YAAD09] E. Yaacoub, H. Al-Asadi, and Z. Dawy. Low complexity scheduling algorithms for the LTE uplink. In *IEEE Symposium on Computers and Communications (ISCC)*, pages 11 – 15, 2009.

[ZHWW07] L. Zhe, Z. He, W. Wu, and X. Wang. A Simplified Layered QoS Scheduling Scheme in OFDM Networks. In *IEEE 66th Vehicular Technology Conference (VTC)*, pages 1842 –1846, September 2007.

[ZXLH08] B. Zhu, K. Xue, H. Lu, and P. Hong. Fair connection admission control scheme for ieee 802.16e systems. In *4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, pages 1 –4, 2008.