

Western  Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

9-16-2013 12:00 AM

Resource Allocation in Uplink Long Term Evolution

Aidin Reyhanimasoleh
The University of Western Ontario

Supervisor
Dr. Abdallah Shami
The University of Western Ontario

Graduate Program in Electrical and Computer Engineering
A thesis submitted in partial fulfillment of the requirements for the degree in Master of
Engineering Science
© Aidin Reyhanimasoleh 2013

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Systems and Communications Commons](#)

Recommended Citation

Reyhanimasoleh, Aidin, "Resource Allocation in Uplink Long Term Evolution" (2013). *Electronic Thesis and Dissertation Repository*. 1627.
<https://ir.lib.uwo.ca/etd/1627>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

RESOURCE ALLOCATION IN UPLINK LONG TERM EVOLUTION
(Thesis format: Monograph)

by

Aidin ReyhaniMasoleh

Graduate Program in Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Engineering Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Aidin ReyhaniMasoleh 2013

Abstract

One of the crucial goals of future cellular systems is to minimize the transmission power while increasing the system performance. This thesis presents two channel-queue-aware scheduling schemes to allocate frequency channels among the active users in uplink LTE. Transmission power, packet delays and data rates are three of the most important criteria critically affecting the resource allocation designs. In the first algorithm, a resource allocation scheme is proposed which aims at minimizing system packet delays of three different data types (video, voice and data) and system transmission power at the same time. After formulating power consumption and packet delays, the four objective functions are collated into a single objective function by using the sum weighting method. propose a way to determine the weights of each objective function and solve the problem by using Binary Integer Programming (BIP). In the second work, we first develop an energy efficient rate adaptive scheduling approach that assigns sub-channels, transport block size, modulation and coding schemes as well as power to the active users in the uplink LTE. The objective function is to maximize the overall throughput of all the active users with respect to uplink standard restrictions and power threshold, which is adaptive at any given frame, per user. Another goal is to guarantee the users' QoS requirements. In the proposed algorithm, we present an approach to reduce power consumption that adjusts the user maximum transmission power threshold according to the QoS requirement for each user. Secondly, due to the high complexity of the adaptive algorithm, a less-complex heuristic algorithm is proposed. The numerical results prove that the adaptive and heuristic algorithms substantially improve the system performance in terms of transmission power while maintaining the demanding users' QoS. In both of the proposed algorithms, the contiguity constraint, which makes the scheduling problem more complicated in uplink rather than downlink is considered.

Acknowledgements

I would like to first and foremost thank my supervisor, Dr. Abdallah Shami, not only for enthusiastically discussing my work with me, answering my questions and giving me good advice, but also for his patience and brilliant personality.

I would also like to thank my parents because without them I would not be here, my siblings (Arash, Azita and Aida) and my uncle (Hooshang) for all of their support. I cannot find a proper word to express my gratitude to them.

Thank you also to everyone in my Lab for all of your scientific and spiritual help.

Contents

Certificate of Examination	ii
Acknowledgements	ii
Abstract	iii
List of Figures	vii
List of Tables	viii
Abbreviations	ix
1 Introduction	1
1.1 Contributions	3
1.2 Thesis Outline	3
2 Long Term Evolution standard overview	4
2.1 System Performance Requirements	4
2.2 Targets for the Long Term Evolution	5
2.2.1 Maximum data rate and spectral efficiency per user	5
2.2.2 Cell throughput and spectral efficiency	5
2.2.3 Mobility	5
2.2.4 User and control plane latency	6
2.2.5 Other parameters	6
2.3 Network Structure	6
2.4 Protocol Architecture	7
2.5 Physical Layer	9
2.5.1 OFDM/OFDMA/SC-FDMA	9
2.5.2 Radio Frame Structures	10
2.5.3 Frequency Domain Organization	12
2.6 Radio Resource Management	14
2.6.1 Admission Control	14
2.6.2 ARQ and HARQ	15
2.6.3 Downlink Dynamic Scheduling and Link Adaptation	16
2.6.4 Uplink Dynamic Scheduling and Link Adaptation	18
2.6.5 Channel State Information	19
2.6.6 Adaptive Modulation and Coding	19

2.6.7	Power Control	21
2.6.8	Buffer Status Reporting	22
3	Resource Allocation in Uplink LTE	23
3.1	Resource Allocation Definition	23
3.2	Resource Allocation Modelling	25
3.3	Search-based Scheduling Models	27
3.3.1	Matrix-based Algorithms	27
3.3.2	Pattern-based Algorithms	28
3.4	The Size of Search Space	30
3.4.1	Scenario 1: assignment of the whole PRBs among users	30
3.4.2	Scenario 2: assignment of all or some PRBs among users	31
3.5	Scheduling Strategy	32
3.5.1	Channel-unaware	32
3.5.2	Channel-aware/QoS-unaware	34
3.5.3	Channel-aware/QoS-aware	36
3.5.4	Power-aware	38
3.6	Literature Review in Uplink LTE Scheduling	38
4	Heterogeneous Delay-Power Resource Allocation in Uplink LTE	48
4.1	Queueing Theory Basics	49
4.1.1	Little's Law	50
4.2	System Model	50
4.2.1	Power criteria	52
4.2.2	Packet delay criteria	54
4.3	Problem Formulation	55
4.3.1	Objective functions and variables	55
4.3.2	Constraints	56
4.4	Resource Allocation Solution	57
4.5	Simulation and Numerical Results	61
4.6	Chapter Summary	65
5	Adaptive Power-efficient scheduler for LTE Uplink	66
5.1	Introduction	66
5.2	System Model	68
5.3	Adaptive Power-Efficient scheduling	69
5.3.1	Delay analysis	70
5.3.2	Adaptive MATP Design and The Objective Function	71
5.4	Heuristic algorithm	72
5.4.1	Complexity of the Heuristic Algorithm	72
5.5	Numerical evaluation	73
5.6	Chapter Summary	75
6	Conclusion and Future Works	77
6.1	Future works	78

Bibliography	79
A Source code of chapter 4	84
B Source code of chapter 5	89
Curriculum Vitae	94

List of Figures

2.1	Overall EPS architecture	7
2.2	LTE Protocol Architecture	8
2.3	Physical Structure in LTE	11
2.4	Frame Structure type 2	12
2.5	Schematic of Downlink Scheduling.	17
2.6	The effect of contiguity constraint on FDPS	19
2.7	Overall view of uplink scheduling in LTE	20
3.1	TDPS/FDPS model	27
3.2	Associated tree for given example.Thick line indicates the assignment	40
3.3	A sample metric matrix	40
3.4	Allocation difference between uplink and downlink	41
3.5	Carrier by carrier method	41
3.6	Bad example of carrier by carrier method	42
3.7	Largest-metric-value-PRB-first method	42
3.8	Bad example of largest-metric-value-PRB-first method	43
3.9	Riding peaks method	43
3.10	Bad example of riding peaks method	44
3.11	The drawback of riding peaks method [1]	45
3.12	PRB grouping method	45
4.1	Queue model	49
4.2	System schematic	51
4.3	PDF of packet delay - one user in the cell	62
4.4	PDF of packet delay - two users in the cell	63
4.5	Average packet delay for different values of MTBS	63
4.6	Average power delay for different values of MTBS	64
4.7	Measure of complexity vs. MTBS	64
4.8	Average power and packet delay vs. No. of users	65
5.1	Average Delay	74
5.2	Average Rates	74
5.3	Normalized Average Power	75
5.4	Normalized Average Time Consumptions	75

List of Tables

2.1	downlink-uplink frame configuration in LTE	12
2.2	Scalable Channel Bandwidth	13
2.3	QCI Characteristics	15
2.4	Supported MCS in LTE	21
3.1	UE-PRB metric matrix	28
3.2	A sample UE-PRB allocation	31
3.3	Example of UE-PRB metric matrix	39
4.1	Summary of notations	52
4.2	Least-Squares Approximate Model Parameters for BLER=10%	54
5.1	List of MCS Indices	69
5.2	Heuristic Allocation	72
5.3	Parameter settings of the uplink LTE model	73

Abbreviations

3GPP	<i>Third Generation Partnership Project</i>
AMC	<i>Adaptive Modulation and Coding</i>
BET	<i>Blind Equal Throughput</i>
BIP	<i>Binary Integer Programming</i>
BSR	<i>Buffer Status Reporting</i>
CDMA	<i>Code Division Multiple Access</i>
CQI	<i>Channel Quality Indicator</i>
CSI	<i>Channel State Information</i>
EDF	<i>Earliest Deadline First</i>
EESM	<i>Exponential Effective SNR Mapping</i>
FDPS	<i>Frequency Domain Packet Scheduling</i>
FIFO	<i>First In First Out</i>
GBR	<i>Guaranteed Bit Rate</i>
HARQ	<i>Hybrid Automatic Repeat reQuest</i>
LTE	<i>Long Term Evolution</i>
LWDF	<i>Largest Weighted Delay First</i>
MAC	<i>Medium Access Control</i>
MCS	<i>Modulation and Coding Scheme</i>
MIESM	<i>Mutual Information Effective SNR Mapping</i>
MME	<i>Mobility Management Entity</i>
MMSE	<i>Minimum Mean Square Error</i>
MT	<i>Maximum Throughput</i>
OFDM	<i>Orthogonal Frequency Division Multiplexing</i>
PAPR	<i>Peak to Average Power Ratio</i>
PF	<i>Proportional Fair</i>
PRB	<i>Physical Resource Block</i>
QoS	<i>Quality of Service</i>
QSI	<i>Queue State Information</i>
RBG	<i>Radio Bearer Group</i>
RR	<i>Round Robin</i>
SC-FDMA	<i>Single-Carrier Frequency Division Multiple Access</i>
SDU	<i>Service Data Unit</i>
SRS	<i>Sound Reference Signal</i>

Chapter 1

Introduction

The development of wireless communication systems has been non-stop in the past decade. First generation cellular networks (1G) were analog-based and limited to voice services only. The first 1G cellular mobile communication system was the Advanced Mobile Phone System (AMPS) that was developed by Bell Labs in the late 1970s [2] and used commercially in the United States in 1983. While these 1G systems give reasonably good voice quality, they offer low spectral efficiency.

This is why the evolution toward 2G was necessary to overcome the drawbacks of 1G technology. The main design objective in Second Generation (2G) cellular networks was to increase voice quality. The second generation of cellular systems, first deployed in the early 1990s, was based on digital communications. The two main categories of 2G cellular systems are GSM (Global System for Mobile Communications) and CDMA (Code Division Multiple Access). The most significant features of GSM that differ from 1G are: (1) using digital cellular technology and (2) exploiting the Time Division Multiple Access (TDMA) transmission method. In the US, 2G cellular networks use direct-sequence CDMA technology with phase shift-keyed modulation and coding. There are three sophisticated versions of GSM [3]:

- High Speed Circuit Switched Data (HSCSD): which yields higher data rates for circuit-switched services as a result of a changing coding scheme and using multiple time slots.
- General Packet Radio Service (GPRS): which had efficient support for non real time packet data traffic. Maximum peak data rates of GPRS are 140 Kbps.

- Enhanced Data rates for Global Evolution (EDGE): which has the maximum data rate 384 Kbps by employing a high-level modulation and coding scheme.

Further progress on the GSM-based and CDMA-based systems have been handled under 3GPP and 3GPP2, respectively. 3GPP introduced the Universal Mobile Telecommunications System (UMTS) as the first global third generation cellular network. The main components of this system are the UMTS Terrestrial Radio Access Network (UTRAN) where Wideband Code Division Multiple Access (WCDMA) radio technology is employed due to its 5 MHz bandwidth, and the GSM/EDGE radio access network based on GSM-enhanced data rates [4]. The third generation continues its improvements and so 3GPP has introduced High-Speed Downlink Packet Access (HSDPA) that resulted in higher speed data services in 2001. Then in 2005, High-Speed Uplink Packet Access (HSUPA) was introduced. The combination of HSDPA and HSUPA is called HSPA [5]. The last evolution of the HSPA category was the HSPA+, which has features such as Multiple Input/Multiple Output (MIMO) antenna capability and 16 QAM (uplink)/64 QAM (downlink) modulation. Due to improvements in the radio access network for packets, HSPA+ will allow speeds of 11 Mbps and 42 Mbps for uplink and downlink, respectively. One of the new concepts in HSPA+ is combining multiple cells into one with a technique known as Dual-Cell HSDPA.

4G networks are sophisticated IP solutions that provide voice, data, and video to mobile users. They offer significantly improved data rates compared with previous generations of wireless technology. Faster wireless connections enable wireless devices to support higher level data services, such as streamed audio and video, video conferencing, gaming and navigation.

As a step toward 4G wireless mobile systems, the 3GPP group began its initial investigation of the Long Term Evolution (LTE) standard in 2004 [6]. Within the 3GPP progress, three multiple access technologies are deployed: the Second Generation including GSM/GPRS/EDGE was based on Time- and Frequency-Division Multiple Access (TDMA/FDMA); the Third Generation UMTS family used Wideband Code Division Multiple Access (WCDMA); finally, LTE has adopted Orthogonal Frequency-Division Multiplexing (OFDM) [7]. LTE mobile systems have been developed by the Third Generation Partnership Project (3GPP) and adopted by the European Telecommunications Standards Institute (ETSI). The finalized technical specifica-

tions of LTE equipment were released (Release 8) at the end of 2008. However, some small enhancements were introduced in Release 9, a release that was functionally finalized in December 2009.

1.1 Contributions

The two main contributions of this thesis are:

In the first algorithm, a resource allocation method which includes packet delays and transmitted power consumption simultaneously is proposed. In this work, after formulating the objective functions (packet delays and power consumption) the scheduler uses the weighted sum method to convert multiple objective functions into a single one. Finally a binary integer optimization method is employed to solve the scheduling problem.

In the second algorithm, a power threshold mechanism is introduced. This power threshold mechanism adapts power threshold for each frame based on the user's Quality of Service (QoS) requirements. The required QoS is power outage delay which implies that probability of outage delay should be less than 2%. In other words, for users who are demanding high QoS, the scheduler increase their power threshold to meet their QoS requirements and the scheduler for users who have low traffic loads decrease their power threshold to save power.

1.2 Thesis Outline

The main objectives of this research are to develop a framework for scheduling to optimally allocate channel, data rate and power resources to multiple users. The remainder of this thesis is organized as follows. Chapter 2 provides a background of LTE Specifications and features. In Chapter 3, the resource allocation problem is surveyed and different types of metrics are explained. Chapter 4 addresses a new heterogeneous delay-power resource allocation in uplink LTE. Chapter 5 states an adaptive power-efficient scheduler for uplink LTE. The conclusion of this thesis and future research suggestions are detailed in Chapter 6.

Chapter 2

Long Term Evolution standard overview

This chapter provides preliminary system information on different specifications of the LTE system. At first, system performance requirements and targets for LTE are presented. The discussion is followed by the network structure and protocol architecture of LTE. Then some aspects of the physical layer in LTE are clarified. Finally, the chapter concludes with the radio resource management concept in LTE with a focus on the scheduling process.

2.1 System Performance Requirements

Before standardization of LTE, 3GPP highlighted the most basic requirements for the LTE:

- The LTE system should be packet switched optimized
- A true global roaming technology with the inter system mobility with GSM, WCDMA and CDMA2000
- Reduced latency with radio round trip time below 10 ms and access time below 300 ms
- Scalable bandwidth from 1.4 MHz to 20 MHz
- Increased spectral efficiency and user data rates
- Simple protocol architecture
- Rational power consumption for the cell phones

- Increased cell-edge bit-rate

2.2 Targets for the Long Term Evolution

The following list is some of the important targets of LTE.

2.2.1 Maximum data rate and spectral efficiency per user

The most important parameter by which the different standards compare with each other is the achievable maximum per-user data rate. This peak data rate depends on used bandwidth and the number of transmitter and receiver antennas in MIMO systems. The maximum data rates for downlink and uplink in the LTE system were set at 100 Mbps and 50 Mbps respectively by using a 20 MHz bandwidth, with the assumption of two receiver antennas and one transmitter antenna for each terminal. Hence maximum spectral efficiencies of 5 and 2.5 bps/Hz are achieved in downlink and uplink LTE, respectively.

2.2.2 Cell throughput and spectral efficiency

Performance at cell level critically depends on the number of cell sites that a network operator needs and therefore determine the main cost of developing a new system. To access the performance at the cell level, 2 metrics are defined: (1) average cell spectral efficiency which is around 1.6-2.1 bps/Hz/cell, (2) cell-edge user spectral efficiency (which used to assess 5% of user throughput) is about 0.04-0.06 bps/Hz/user.

2.2.3 Mobility

In terms of mobility, the LTE network supports communication with terminals moving at speeds of up to 350 km/h according to the speed of future trains. Due to this high speed, the complexity of the LTE system is increased so that the handover between cells is done without interruption.

2.2.4 User and control plane latency

The average time between the sending of a data packet and the reception of a physical layer Acknowledgement (ACK) determines user plane latency (by considering typical HARQ re-transmission rates). Simplicity, the round trip time is twice of user plane latency. Reduction of call set-up delay is one of the significant requirements of LTE system. This results in both good user satisfaction and more importantly affects the battery life of terminals. In other words, control plane latency is the amount of time delay between the sending of a command message or the initiation of a service request to when the command begins to process or the service begins to operate.

2.2.5 Other parameters

Besides the system performance aspects, a number of other criteria are important for network operators. These include reduced deployment cost, bandwidth flexibility, compatibility with other radio access technologies, and lower power consumption terminals [7].

2.3 Network Structure

The overall architecture of LTE has two distinguished components: the radio access network and the core network. The first one, called Evolved Universal Terrestrial Radio Access Network (E-UTRAN), and the second one, which is fully Packet Switched is called System Architecture Evolution (SAE). E-UTRAN compared to its ancestors such as UMTS, HSDPA and HSUPA is completely different and resulted in higher data rates and lower latencies. E-UTRAN consists of two components: (1) access point eNodeB (the same as Base Station) which supply one or more cells, and (2) User Equipment (UE). eNodeBs are connected to each other by X2 interface. Inter-cell interference information can be transferred between Base Stations (BSs) over this interface. BSs connect to the core network via the S1 interface. The main component of the SAE architecture is the Evolved Packet Core (EPC) which includes: (1) Serving Gateway (S-GW) which is responsible for handovers with the neighbour eNodeBs (2) Mobility Management Entity (MME) which is responsible for idle mode UE tracking and paging procedures,

and (3) Packet Data Network Gateway (PDN-GW) which is responsible for IP address allocation for the UE, as well as QoS enforcement for Guaranteed Bit Rate (GBR) bearers. Note that E-UTRAN and EPC together constitute the Evolved Packet System (EPS). Figure 2.1 shows the overall EPS architecture.

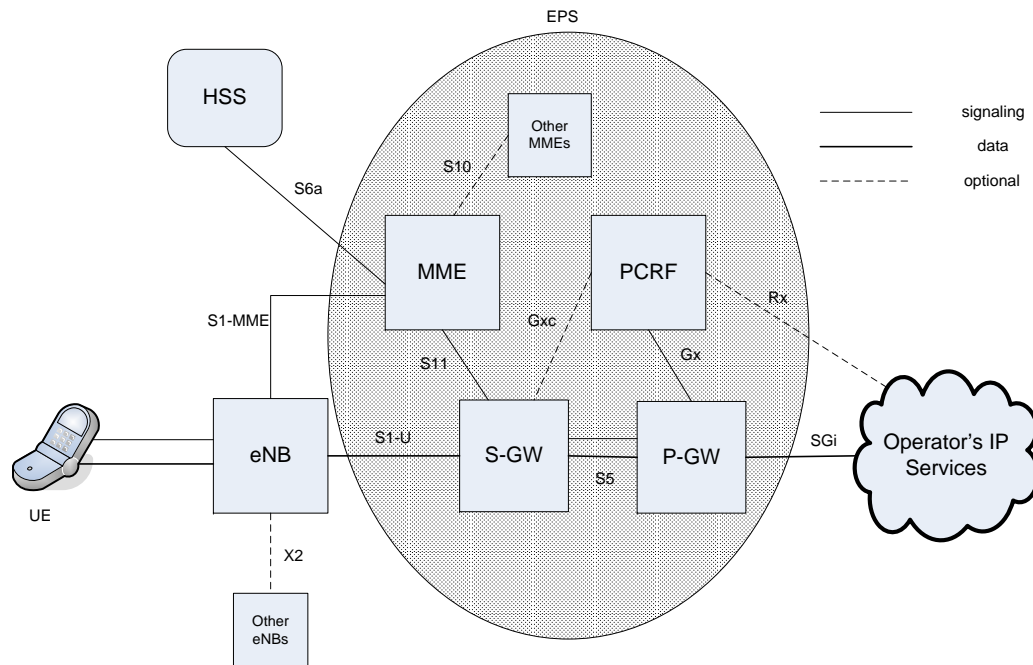


Figure 2.1: Overall EPS architecture

2.4 Protocol Architecture

Radio protocol stack layers in LTE can be divided into three layers which consists of the physical layer (layer 1), the data link layer (layer 2), and the network layer (layer 3). Radio Resource Control (RRC) is the main sub-layer of layer 3. The most important tasks which should be carried out by RRC include making handover decisions concerning neighbour cell measurement sent by UE and setting up radio bearers. Data link layers consists of three sub-layers. At the top, Packet Data Convergence Protocol (PDCP) performs compressing/decompressing of the headers of IP packets using Robust Header Compression (ROHC) and data integrity with enciphering. PDCP hands its packets, namely Service Data Units (SDUs) to the intermediate

sub-layer of the link layer i.e. Radio Link Control (RLC). RLC is responsible for reassembly of SDUs into Protocol Data Units (PDUs). This reassembly can be either segmentation of one SDU into several PDUs or, concatenation of several SDUs to form one PDU due to the transmission data rate. The lowest sub-layer of the link layer is the Medium Access Control (MAC) sub-layer that handles Hybrid Automatic Repeated reQuest (HARQ) functionality and scheduling problems. The Physical (PHY) layer performs all of the tasks regarding the transmission of actual data to air interface such as modulation and coding. Figure 2.2 depicts the protocol architecture of LTE. As shown in the Figure the channel between air and PHY layer is the physical channel, the channel between PHY layer and the MAC sub-layer is the transport channel and between the MAC and the RLC sub-layers is the logical channel.

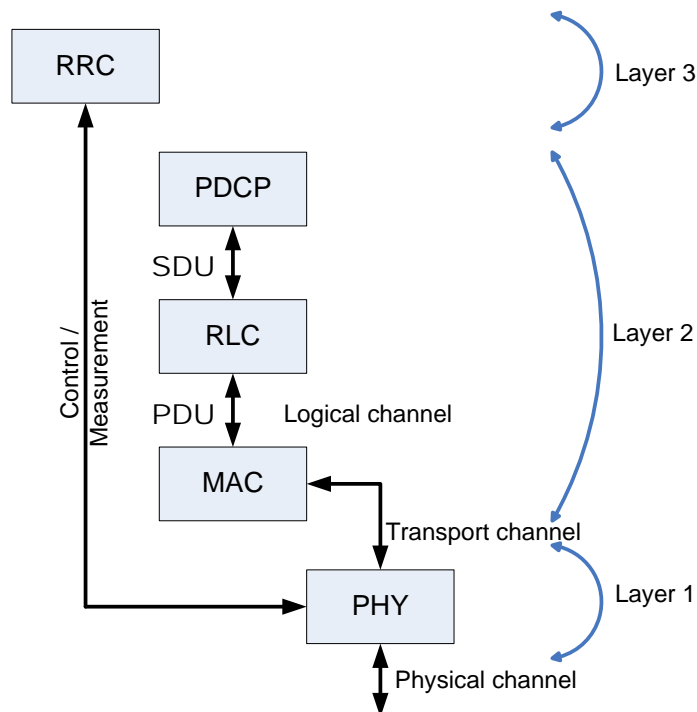


Figure 2.2: LTE Protocol Architecture

2.5 Physical Layer

In this section, different features and specifications of the physical layer in LTE are briefly investigated. A comprehensive investigation of this concept by itself needs several hundred pages to cover them. For the sake of brevity, this section just explores the key concepts of the physical layer.

2.5.1 OFDM/OFDMA/SC-FDMA

Orthogonal Frequency Division Multiplex (OFDM) is a combination of modulation and multiplexing where modulation is a mapping of information on changes in the carrier phase, frequency or amplitude or some combination. Multiplexing is a method of sharing bandwidth with other independent data channels. When using single carrier modulation over a multi-path channel, channel delay spread may be longer than the symbol duration. This situation results in Inter-Symbol Interference (ISI) at the receiver. In order to demodulate the data, a system would have to employ an equalizer to reduce ISI. OFDM overcomes the ISI problem by modulating several narrow-band sub-carriers in parallel. Since any sub-carrier has a narrow bandwidth, it is not influenced from block fading and hence does not experience ISI. Not all the sub-carriers carry data. For synchronization purposes, some of the sub-carriers, namely pilot-carriers, are modulated with a constant pattern known to both the transmitter and the receiver. Also some sub-carriers at the edges of the frequency band are not modulated, and serve as a guard band. The other modulated sub-carriers are multiplexed using an Inverse Fourier Transform (IFFT). A cyclic prefix is added to the resulting time-domain waveform. OFDMA is a combination of a modulation scheme the same as OFDM and a multiple access scheme that combines TDMA and FDMA. In OFDM, at each given time, only one user can transmit on the entire bandwidth, and so time division multiple access is employed to serve multiple users. OFDM ensures that sub-carriers are assigned to the users have good channels because the OFDMA allows several users to send data on the different sub-carriers per OFDM symbol at the same time. In other words, OFDM allocates users in the time domain only, but OFDMA allocates users in time and frequency domains. As a result, OFDMA is adopted as the downlink multiple access scheme in LTE by 3GPP due to following advantageous:

- High spectral efficiency, which is also called bandwidth efficiency. This term means that more data can be transmitted in presence of the noise in a given bandwidth during a fixed time interval. The unit of spectral efficiency is bits per second per Hertz (b/s/Hz).
- Robustness to multi path delay spread as a result of long symbol time and guard interval
- Flexible utilization of frequency spectrum
- Low-complexity receivers, by exploiting frequency-domain equalization
- Effectiveness against Channel Distortion due to utilization narrow bandwidth

However, despite its many advantages, OFDMA has certain drawbacks such as high sensitivity to frequency offset and high peak-to-average power ratio (PAPR) due to in-phase addition of subcarriers. The second one plays an important role in uplink direction where transmitters are cell phones with limited power storage. 3GPP selected Single Carrier FDMA (SC-FDMA) as multiple access method of uplink in LTE because this scheme possesses the same advantages of OFDMA while experiencing lower PAPR. The adoption of SC-FDMA enhances the power consumption efficiency of the cell phone batteries, hence prolonging their lifetimes. SC-FDMA exhibits 3-6 dB less PAPR than OFDMA. The main reason for the selection of SC-FDMA among other PAPR reduction methods is the similarity of SC-FDMA to OFDMA in implementation structure. Figure 2.3 illustrates the structure of the physical layer in LTE. As can be seen, the only difference between SC-FDMA and OFDMA is the presence of a DFT and an IDFT block in the transmitter and receiver, respectively. That is why that SC-FDMA is also known as DFT pre-coded OFDMA.

2.5.2 Radio Frame Structures

Release 9 LTE includes two types of frame structures: (1) Type 1, which uses Frequency Division Duplexing (FDD) and (2) Type 2, which uses Time Division Duplexing (TDD). In Type 1, downlink and uplink transmissions employ different frequency bands and each has its own frame. In this type, the radio frame has a 10 ms duration, which is divided into 10 subframes (each being 1ms long). Sub-frames are the fundamental time unit for most LTE

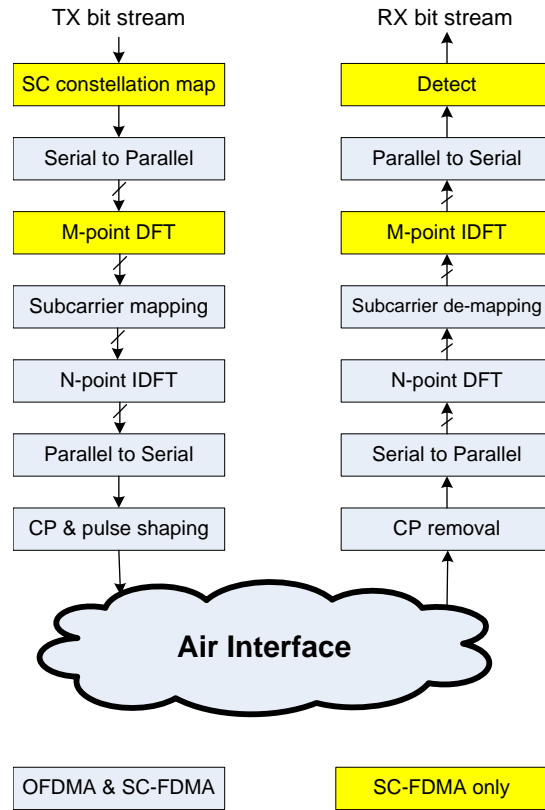


Figure 2.3: Physical Structure in LTE

processing, like scheduling. Each sub-frame consists of two time slots, which are each 0.5ms long. Each time slot depends on the duration of Cyclic Prefix (CP), which has 6 or 7 OFDM/SC-FDMA symbols. Frame structure Type 2 is only applicable to TDD, which utilizes the same frequency band in uplink and downlink and shares frames in time domain. The structure of each Type 2 frame is identical to Type 1. The only difference is the existence of one or two special sub-frames that help switching between uplink and downlink transmissions. These special sub-frames have three special fields: the downlink pilot timeslot (DwPTS), the guard period (GP) and the uplink pilot timeslot (UpPTS). The duration of these three fields is equal to one sub-frame. Figure 2.4 shows the Type 2 frame structure of LTE.

There are different downlink-uplink frame configurations in LTE as illustrated in Table 2.1. In this Table, D and U are respectively downlink and uplink transmissions, while S is

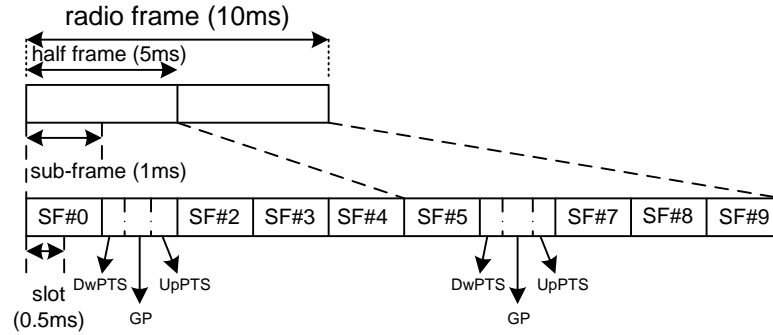


Figure 2.4: Frame Structure type 2

a special sub-frame for the switching purpose. Note the sub-frame 0 and sub-frame 5 in all configurations are for downlink. Sub-frames immediately following the special sub-frame (i.e., sub-frame 2 in all configurations and sub-frame 7 in 5ms periodicity) are always reserved for the UL transmission.

Configuration #	DL to UL switch priority	sub-frame number									
		1	2	3	4	5	6	7	8	9	10
0	5ms	D	S	U	U	U	D	S	U	U	U
1	5ms	D	S	U	U	D	D	S	U	U	D
2	5ms	D	S	U	D	D	D	S	U	D	D
3	10ms	D	S	U	U	U	D	D	D	D	D
4	10ms	D	S	U	U	D	D	D	D	D	D
5	10ms	D	S	U	D	D	D	D	D	D	D
6	5ms	D	S	U	U	U	D	S	U	U	D

Table 2.1: downlink-uplink frame configuration in LTE

2.5.3 Frequency Domain Organization

LTE DL/UL air interface waveforms use several orthogonal subcarriers to send user traffic data, reference signals (pilots), and control information. The frequency spacing between subcarriers is 15KHz. The smallest modulation structure in LTE is the Resource Element. A

Resource Element is one 15 kHz subcarrier by one symbol. Resource Elements aggregate into Resource Blocks. A Resource Block has dimensions of subcarriers by symbols. Twelve consecutive subcarriers in the frequency domain and six or seven symbols in the time domain form each Resource Block. As noted, the number of symbols depends on the Cyclic Prefix (CP) in use. When a normal CP is used, the Resource Block contains seven symbols. When an extended CP is used, the Resource Block contains six symbols. A delay spread that exceeds the normal CP length indicates the use of extended CP. Various channel bandwidths that may be considered for LTE deployment are shown in Table 2.2.

Channel Bandwidth (MHz)	1.4	3	5	10	15	20
No. of Sub-carriers	73	181	301	601	901	1201
FFT Size	128	256	512	1024	1536	2048
Sampling Rate (MHz)	1.92	3.84	7.68	15.36	23.04	30.72
No. of PRBs	6	15	25	50	75	100

Table 2.2: Scalable Channel Bandwidth

UL/DL resource grid definitions are summarized as:

- Resource Element (RE): One element in the time/frequency resource grid. One sub-carrier in one OFDM/SC-FDMA symbol for DL/UL. Often used for Control channel resource assignment.
- Physical Resource Block (PRB): 12 consecutive sub-carriers (180 kHz) over the duration of one slot, which is the minimum scheduling size for DL/UL data channels.
- Resource Block Group (RBG): Group of Resource Blocks where the size of RBG depends on the system bandwidth in the cell.
- Resource Element Group (REG): Groups of Resource Elements to carry control information. The size of REG is four or six REs depending on the number of reference signals per symbol, cyclic prefix length.
- Control Channel Element (CCE): Group of nine REGs form a single CCE and are used for control information. Both REG and CCE are used to specify resources for LTE DL control channels.

2.6 Radio Resource Management

The aim of RRM is to maximize the radio resource efficiency by utilization of the adaptation techniques and satisfying the configured users' Quality of Services. There are two categories of RRM algorithms: (1) semi-dynamic category; which are mainly executed during the setup of new data flows and (2) fast dynamic category named such since every action is carried out at each sub-frame (1 ms). The semi-dynamic category consists of three algorithms: QoS management, admission control, and semi-persistent scheduling, all of which are in Layer 3. The fast dynamic category includes Hybrid Adaptive Repeat and Request (HARQ) management, dynamic packet scheduling, and link adaptation in Layer 2 as well as the Channel Quality Indicator (CQI) manager, and power control in Layer 1.

2.6.1 Admission Control

The task of admission control is to accept or reject the requests of new Evolved Packet System (EPS) bearers in the cell. This decision is made according to the available resources of the cell, the QoS provisions for the new EPS bearer and the provided QoS to the active users in the cell. A new request is only accepted if the algorithm predict that the following conditions are satisfied: (1) QoS for the new EPS bearer can be met, and (2) promised QoS requirements are fulfilled for all the existing bearers in the cell with the same or higher priority level. It is worth noting that the precise decision mechanism and algorithms for admission control are not determined by 3GPP and it is eNB vendor-based. Each LTE EPS bearer has its own QoS specifications. All the packets within the bearer have the same QoS parameters. The information that are associated with the QoS profile of the EPS bearer include: (1) allocation retention priority (ARP), (2) uplink and downlink guaranteed bit rate (GBR), and (3) QoS class identifier (QCI). In LTE, there are GBR bearers and non-GBR bearers and GBR parameters only exist for GBR bearers. ARP is an integer between 1 and 16, which represents the priority level of the bearer that is utilized for the admission control mechanism. QCI is a scalar that represents the specifications of the specific bearer (e.g. bearer priority, packet delay budget and packet loss rate), and that have been preconfigured by the operator owning the eNB. Table 2.3 shows nine different QCIs and their typical features defined in LTE standard [8] [9].

QCI #	Type	Priority	Packet delay budget	Packet loss rate	Example services
1	GBR	2	100ms	10^{-2}	Conversational voice
2	GBR	4	150ms	10^{-3}	Conversational video
3	GBR	5	300ms	10^{-6}	Buffered streaming
4	GBR	3	50ms	10^{-3}	Real time gaming
5	non-GBR	1	100ms	10^{-6}	IMS signalling
6	non-GBR	7	100ms	10^{-3}	Live streaming
7	non-GBR	6	300ms	10^{-6}	Buffered streaming, email,
8	non-GBR	8	300ms	10^{-6}	browsing, file download,
9	non-GBR	9	300ms	10^{-6}	file sharing, etc.

Table 2.3: QCI Characteristics

In uplink, per bearer, there is another QoS parameter named prioritized bit rate (PBR). The aim of PBR is to avoid uplink scheduling starvation problems for UEs with multiple bearers. PBR differs from GBR and can also be defined for non-GBR bearers. The uplink rate control mechanism ensures that the UE serves at first the radio bearers in decreasing priority order up to their PBR, and then the radio bearers in decreasing priority order for the remaining resources.

2.6.2 ARQ and HARQ

As in any communication system, there are data transmission errors, which can be due to noise and interference. Most of the protocols are not able to correct errors in the data packets. To solve this problem, complementary mechanisms are required. An approach is to deploy backward error correction (aka Automatic Repeat Request). In ARQ, the receiver informs the transmitter whether a data packet was received correctly or not. If the reception is erroneous, the transmission is repeated. Although this mechanism is simple and significantly efficient, there are some drawbacks as listed below:

- ARQ results in delay in transmission of data packets and this delay is grown out of feedback response and retransmission if data are transmitted incorrectly.
- ARQ is efficient if the average packet error rate is reasonably small.
- The feedback loop has to be protected against errors.

ARQ is not optimal because it throws away the information in the erroneous packet. A superior method is that the receiver stores and exploits all of the past received information. Even if the received data from the first transmission is not enough for successful decoding, it can still be helpful if combined with the second transmission. This scheme is called Hybrid ARQ (HARQ). In general, HARQ schemes can be categorized as adaptive-synchronous, non-adaptive-synchronous, adaptive-asynchronous and non-adaptive-asynchronous. In a synchronous HARQ schemes, the retransmission time relative to the first transmission is specified and so there is no need for an information signal, for example a HARQ process number. However, in an asynchronous HARQ scheme, the retransmissions can happen at any time after the first transmission, which causes asynchronous HARQ to need extra signalling to transmit the HARQ process number to the receiver. As a result, synchronous HARQ schemes have the advantage of decreasing the signalling load and the disadvantage of less flexibility in scheduling compared to asynchronous HARQ schemes. In an adaptive HARQ scheme, the retransmissions can be employed either the same or with another modulation and coding scheme and resource allocation in the frequency domain relative to initial transmission. The changes in transmission attributes arises from variation in the channel condition. This means this scheme needs additional signalling. By contrast, in the non-adaptive HARQ scheme, the retransmissions do not need the explicit signalling of new transmission attributes, since retransmissions are executed either the same as the initial transmission or with new attributes, which is determined according to a predefined regulation. In summary, adaptive schemes have more scheduling gain at the expense of increased signalling overhead. In LTE, asynchronous adaptive HARQ is used for the downlink, and synchronous HARQ for the uplink. In the uplink, the retransmissions may be either adaptive or non-adaptive depending on whether new signalling of the transmission attributes is provided.

2.6.3 Downlink Dynamic Scheduling and Link Adaptation

A dynamic scheduler entity in layer 2 performs packet scheduling to achieve high spectral efficiency while meeting the required QoS in the cell. The scheduling decisions are made every TTI and scheduler functionality is to allocate Physical Resource Blocks to the users, as well as

transmission parameters such as modulation and coding schemes, which is called link adaptation. In other words, the purpose of packet scheduling is to maximize the cell throughput, while the minimum QoS requirements for the EPS bearers are met and remain adequate resources for best-effort bearers. The best-effort bearers have no strict QoS requirements. Figure 2.5 depicts the general schematic of downlink scheduling.

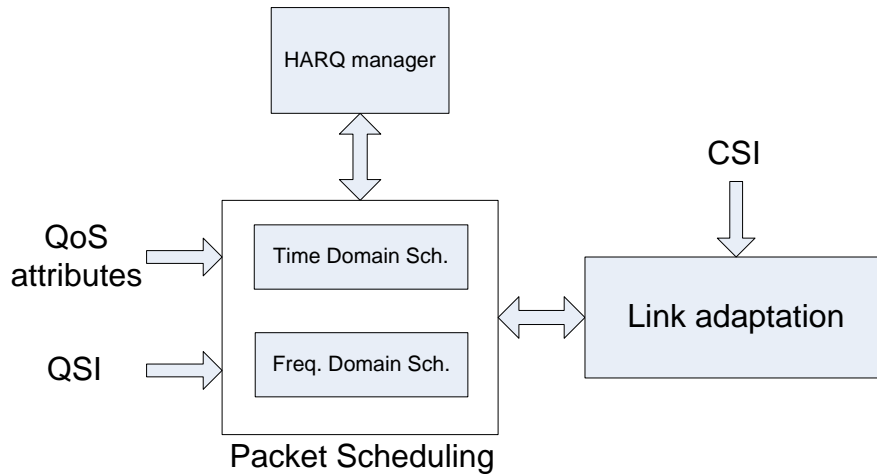


Figure 2.5: Schematic of Downlink Scheduling.

The scheduling decisions are made per user and each user can have multiple data flows. The packet scheduler communicates with the HARQ manager for scheduling retransmissions. As noted, in LTE asynchronous adaptive HARQ is used for the downlink and for each TTI scheduler must send either a new transmission or a pending HARQ retransmission for each individual user and cannot send both of them. The link adaptation gives some information about the supported modulation and coding scheme for a user to the packet scheduler. The link adaptation unit infers this information from the users' CQI feedbacks in the cell. Frequency Domain Packet Scheduling (FDPS) is a striking method to improve the LTE system throughput. This method utilizes frequency selective fading of signal. In other words, the scheduler assigns PRBs to the users that experienced the higher channel quality. In LTE, time domain scheduling gain is low due to using relatively large bandwidth and multiple antennas. Another scheduling method is joint time and frequency domain scheduling. In this method, at first, time scheduling selects N users according to the associated priority metric and passes these users to

the frequency domain scheduler and then the frequency domain scheduler assigns PRBs to the selected users. The complexity of this method is much lower than fully time/frequency domain scheduler, while it has almost the same performance. If HARQ retransmissions are included in scheduling, Time Domain Scheduler (TDS) passed all of the users that have pending HARQ retransmission to FDPS. Two scenarios exist for HARQ-aware FDPS. In scenario #1, in the first step, N_{harq} PRBs are reserved for all of the users with pending HARQ retransmissions. In the second step, all of the remaining PRBs are assigned to the users with new data packets based on FDPS metric value (this metric depends on many parameters such as channel gain and so on) and in the third step, the remaining PRBs are allocated to HARQ retransmissions. Scenario #2 is the same as #1 but exchanges the order of the second and third steps. In both scenarios, it is assumed the number of required PRBs for retransmission is the same as the initial transmission. It is obvious that the latter scenario gives the higher priority to the HARQ retransmission relative to the former scenario.

2.6.4 Uplink Dynamic Scheduling and Link Adaptation

In uplink LTE, there are some special features that make the scheduling in uplink different from that in downlink. The three main differences are listed as follows:

1. The first and the most important distinction is PRB contiguity allocation constraint. Contiguity constraint implies all of the multiple PRBs assigned to a certain user have to be adjacent to each other. This limitation is derived from SC-FDMA. Figure 2.6 illustrates the comparison of uplink/downlink FDPS with/without contiguity constraint. This constraint limits both frequency and multi-user diversity.
2. In uplink, data transmitters are UEs that have limited transmitter power compared to base stations in downlink. On the other hand, UEs tend to decrease power consumption to prolong the battery life time of UEs. In summary, uplink has less power budget relative to downlink.
3. In uplink, eNB does not have complete information of the user's queue size. This feature is explained later in the buffer status report section.

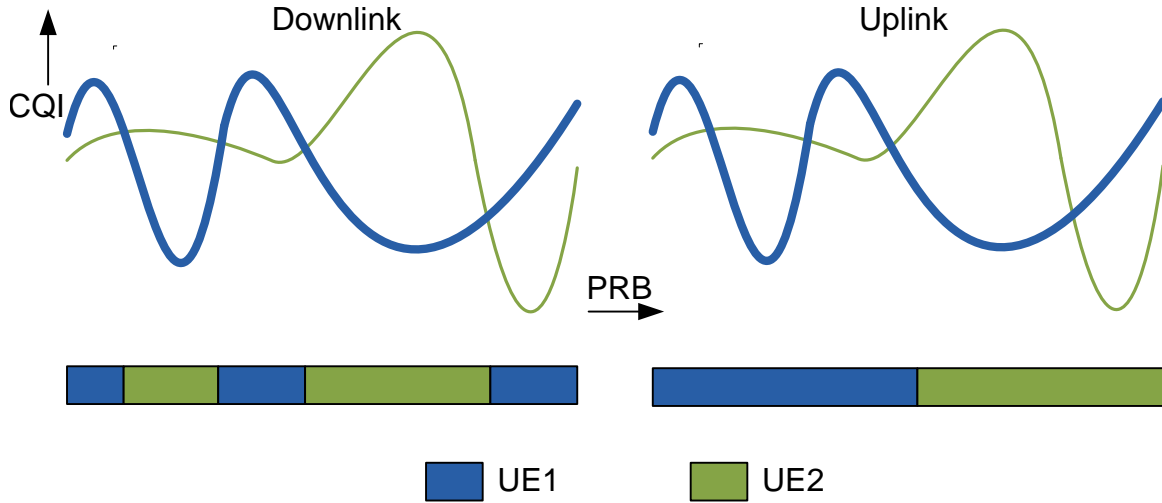


Figure 2.6: The effect of contiguity constraint on FDPS

Figure 2.7 shows the overall view of uplink scheduling in LTE.

High efficient packet scheduling and link adaptation are strongly related to two main categories of information, which are Channel State Information (CSI) and Queue State Information (QSI).

2.6.5 Channel State Information

CSI is applied to AMC block to affect selection of MCS and Scheduling block to perform FDPS. CSI is calculated based on the SNR measurements of Sound Reference Signals (SRSs) in uplink. Allocation of SRS resources among the users is one of the RRM functions in uplink. The purpose of allocation is to update channel state information. There is a compromise between measurement precision and SRS bandwidth in such a way that, by decreasing the SRS bandwidth, the measurement becomes more accurate. However, to know of the entire bandwidth, several SRS transmissions are required.

2.6.6 Adaptive Modulation and Coding

This block has two main tasks. The first task is to report channel state information of the users to the packet scheduler and hence AMC block acts as an interface between the CSI

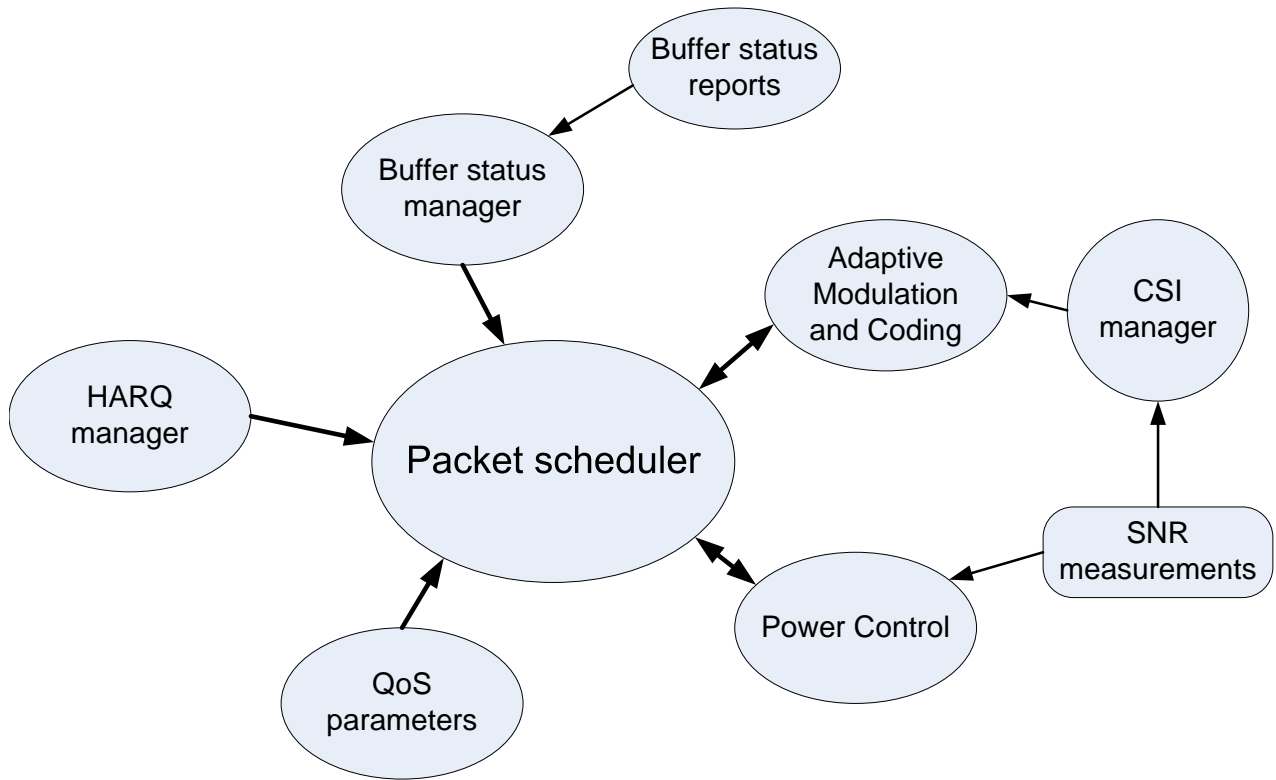


Figure 2.7: Overall view of uplink scheduling in LTE

manager and the packet scheduler. The second task is to select the most efficient MCS for a certain user once the allocated bandwidth for the corresponding user is specified. By using a proper AMC, obviously the spectral efficiency of a wireless system is increased. In practice, AMC is performed by employing AMC mapping tables. These tables return the MCS and the corresponding Transport Block Size based on SNR value and the given Block Error Rate (BLER). At any TTI, AMC selects the MCS that maximizes the expected transport block size (T). Expected transport block size is a function of TBS and Block Error Probability (BLEP), which infers the probability of the erroneous transmitted block as shown in the following:

$$T(MCS, SNR) = TBS(MCS) \times [1 - BLEP(MCS, SNR)] \quad (2.1)$$

In terms of the periodicity of AMC, there are two categories of AMC: (1) slow AMC, in which AMC is done in the slow rate, for example with the same rate of the power control

commands and (2) fast AMC; in which AMC is performed at each TTI. Clearly, the fast AMC leads to better gain compared to the slow one. That is why all of the schedulers select fast AMC as a default. All of the supported MCS in LTE and their characteristics are illustrated in Table 2.4.

Index	Modulation	Code Rate	Spectral Efficiency
0	-	-	-
1	QPSK	78/1024	0.1523
2	QPSK	120/1024	0.2344
3	QPSK	193/1024	0.3770
4	QPSK	308/1024	0.6016
5	QPSK	449/1024	0.8770
6	QPSK	602/1024	1.1758
7	16QAM	378/1024	1.4766
8	16QAM	490/1024	1.9141
9	16QAM	616/1024	2.4063
10	64QAM	466/1024	2.7305
11	64QAM	567/1024	3.3223
12	64QAM	666/1024	3.9023
13	64QAM	772/1024	4.5234
14	64QAM	873/1024	5.1152
15	64QAM	948/1024	5.5547

Table 2.4: Supported MCS in LTE

2.6.7 Power Control

The main goal of power control is to limit inter-cell interference while considering QoS requirements and to minimize UE power consumptions to prolong the battery life of users. Based on [10], transmit power of each UE can be calculated by the following equation:

$$P = \min\{P_{max}, P_0 + 10\log_{10}N + \alpha L + \Delta_{MCS} + f(\Delta_i)\} \quad (2.2)$$

where P_{max} is the maximum user transmission power, N is the number of allocated PRB at a given TTI, P_0 and α are power control parameters, L is the downlink path-loss measured in the UE and is a function of distance, path loss, shadowing and antenna gain. Δ_{MCS} is a cell dependent factor given by Radio Resource Control (RRC), $f(\Delta_i)$ is a user specific closed loop

correction. It is noticed that α is cell-dependent and takes the value zero or 0.4 to 1.0 with the step of 0.1, while P_0 can either be cell- or user- dependent. As a result, the task of power control is to (1) modify the transmission power of users with respect to radio propagation channel, including path loss, shadowing and fast fading and (2) overcome interference from inter-cell and intra-cell users.

2.6.8 Buffer Status Reporting

In LTE, Buffer Status Reporting (BSR) includes the buffer size of several Radio Bearer Groups (RBGs) for each user. This scheme offers relatively low signalling load and high flexibility in scheduling. BSR consists of at most four different RBGs to report. The mapping of each radio bearer to the corresponding RBG is performed based on vendor-specific mapping tables by considering the radio bearer QoS. The buffer size of each RBG represents the amount of data relevant to radio bearers of a certain RBG. There are two formats of BSR in LTE.

- Short BSR format: in this format, a certain user just sends the buffer size of one RBG and the identifier of the transmitted RBG.
- Long BSR format: all of the four buffer sizes of each user are transmitted.

The Buffer Status reporting procedure is used to analyze delay and hence to devise superior scheduling algorithms.

Chapter 3

Resource Allocation in Uplink LTE

This chapter provides more detailed discussion of the resource allocation problem. This chapter starts by the defining resource allocation problem and modelling. Then, two types of search-space scheduling models, as well as the largeness of the search space, are explained. Different scheduling strategies are then investigated. Finally, a literature review of existing works regarding resource allocation in uplink LTE has been provided.

3.1 Resource Allocation Definition

In wireless shared bandwidth networks, resource allocation is defined as allocation of a portion of bandwidth and power to different users to improve network performance. In uplink LTE, since different MCSs can be supported, the most efficient MCS should be assigned to the user in addition to physical resource blocks and power. All of the scheduling tasks are performed in the MAC sub-layer located in the eNB. Because of SC-FDMA characteristics, channel variation in space, time and frequency per user can be utilized by the scheduler. A good scheduler should contain two attributes at the same time. The first one is to satisfy the QoS requirements of users and the second one is to increase the efficiency of resources allocated to the users. In general, the scheduler should take into account some or all of the following factors simultaneously as follows:

- CSI: provide the channel quality information between users and eNB over different PRBs. The information is used by the scheduler to efficiently assign the PRBs to users.

- QSI: with knowledge of users' QSI, the scheduler assigns more PRBs to the users which have more available data in their buffers. Also the scheduler ensures not to assign transport block sizes more than the available queue size of each user.
- QoS requirements: the scheduler must guarantee to provide the user's QoS. In the sophisticated schedulers, each user has different traffic types which have their own QoS (such as average delay, guaranteed bit rate and packet error rate).
- HARQ retransmission: the scheduler decides which PRBs should be reserved for HARQ retransmissions and which ones for new transmissions.
- Maximum No. of users: in some of the scheduling algorithms, a predefined maximum No. of users is allowed to be served at each TTI.
- History of user rates: this history can be deployed to consider the fairness of the users. In the channel-aware scheduling, the users close to the edge of the cell experience pretty bad channel quality rather than users close to the eNB and so have a lower chance to take the bandwidth. To avoid this unfairness in taking sub-channels, the history of user rates are included in the scheduling.
- User priority: some users have more priority than others. This priority should be considered in the scheduling problem.
- Allocation constraint: as with other shared bandwidth resource allocation schemes, each PRB can be assigned to at most one user.
- Contiguity constraint: SC-FDMA imposes contiguity constraints in the uplink scheduling. According to this limitation, all of the allocated PRBs to each user have to be adjacent to each other. This restriction makes the scheduling problem more complicated in uplink compared to downlink.
- Uplink transmission power: less transmission power consumption results in longer UE battery life. Hence devising the energy efficient scheduling algorithms is one of the targets in uplink LTE resource allocation.

- Complexity: packet scheduling decisions are made in sub-frame duration (1ms). Thus the scheduling scheme should have low complexity to limit processing time and memory usage.

The scheduling algorithm takes into account some of or all of the above factors to maximize or minimize a desired aim. The most important objectives are listed as follows:

- Maximization of the overall cell throughput: one of the most important performance indicators in effective utilization of radio interface in any cellular network is the actual throughput or spectral efficiency (expressed in bit/s/Hz). Actual throughput refers to data rate without including HARQ retransmissions. The overall cell throughput can be calculated as a summation of active user throughput of the cell.
- Maximization of fairness: a blind maximization of the overall cell throughput leads to an unfair resource sharing among users. If the scheduler just focuses on spectral efficiency, the users with bad channel quality (such as cell-edge users) can have less opportunity to take allocation resources.
- Minimization of power consumption: in uplink, power consumption is an important feature which should be considered in scheduling to prolong the battery life time of cell-phones. Power consumption in uplink is more important than that in downlink because transmitter units in uplink are cell-phones fed from limited energy batteries while in downlink the transmitter units are eNBs with unlimited energy suppliers.
- QoS provisioning: some schedulers just emphasize the satisfaction of users' QoS requirements.

3.2 Resource Allocation Modelling

LTE uplink resource allocation can be considered an optimization problem where objective function represents the desired performance metric and the solution is the mapping resources, especially PRBs, among active users. By considering all of the factors which are introduced in section 3.1 and scanning all of the available patterns for assignment PRBs to users, coming up

with the optimal solution can be complicated. By using this model, each scheduling scheme includes two stages:

1. Determination of the objective function: objective function is a mathematical formula which maps a satisfaction level of the system performance to a quantitative value. Based on the scheduler's strategy, the desired system performance can include one or a combination of the mentioned objectives in Section 3.1. The satisfaction level of the system is related to the satisfaction level of the available users in the cell. Hereafter, quantitative value of the user's satisfaction level will be known as user utility and denotes as U_i and the quantitative value of the system's satisfaction level will be known as system utility and denotes as U_{sys} . Apparently, U_{sys} is a function of U_i and the simplest form of this function is summation.

$$U_{sys} = \sum_{i=1}^K U_i \quad (3.1)$$

Depending on the parameters which are included in the utility function, system utility values differ in each TTI.

2. Determination of the search based allocation scheme: the scheduler runs an algorithm which searches among all of the UE-PRB allocation patterns until it comes up with the pattern which best optimizes the defined system utility function. The scheduler should implement search based allocation algorithms once per sub-frame (1ms). Hence, devising a low complex algorithm that approximates to that of the optimal algorithm is of great importance. According to the selected performance strategy, the optimization problem can be a minimization algorithm (e.g. packet delay or packet loss rate) or maximization one (like system throughput or fairness among users).

A well-known model in the scheduling scheme is to break the scheduling problem into two different blocks with different aims for each one. The first block is Time Domain Packet Scheduling (TDPS) whose aim is to prioritize users and select some of them to be scheduled for the current sub-frame. The second block is Frequency Domain Packet Scheduling (FDPS) whose aim is to select the best UE-PRB mapping in terms of desired system utility function. This model is known as the TDPS/FDPS model. Figure 3.1 illustrates the structure of this model.

The simplified version of this model is that it bypasses the first block. In other words, all of the active users in the cell passed into the FDPS block.

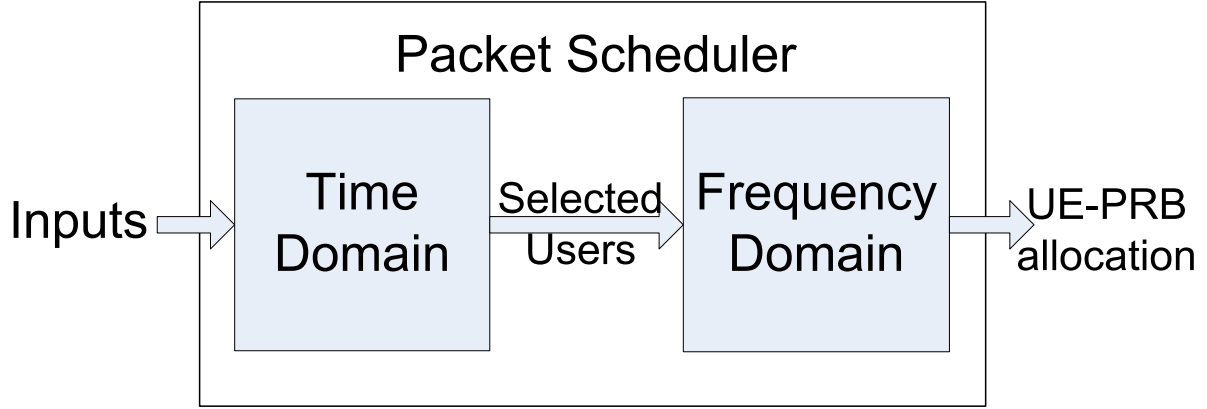


Figure 3.1: TDPS/FDPS model

3.3 Search-based Scheduling Models

As noted, the solution in packet scheduling is to find the best UE-PRB pattern to maximize or minimize system performance utility. Almost all of the search based algorithms in uplink LTE can be classified as one the following models.

3.3.1 Matrix-based Algorithms

In this model, the scheduler forms a matrix. The matrix has K rows according to the number of active users in the cell and M columns relevant to the number of PRBs which can be scheduled. Each element of this matrix represents a metric value that is achieved from the utility function where $M_{i,m}$ denotes the metric value for user i and PRB m . Table 3.1 shows the UE-PRB metric matrix. The allocation algorithms select the maximum metric value in the matrix and assign the corresponding PRB to the associated user by keeping in mind the defined standard constraint (allocation and contiguity constraints).

In the TDPS/FDPS model, in addition to the metric matrix which is related to the FDPS, the scheduler should form a metric vector for the TDPS. The task of this vector is to weigh

	PRB_1	PRB_2	\dots	PRB_M
UE_1	$M_{1,1}$	$M_{1,2}$	\dots	$M_{1,M}$
UE_1	$M_{2,1}$	$M_{2,2}$		$M_{2,M}$
\vdots	\vdots			\vdots
UE_K	$M_{K,1}$	$M_{K,2}$	\dots	$M_{K,M}$

Table 3.1: UE-PRB metric matrix

the importance of users based on the selected policy and then select the set of users with maximum metric to pass into the FDPS block. Therefore, in this model, two utility functions and accordingly two metrics should be defined.

3.3.2 Pattern-based Algorithms

In this model, the scheduler forms one binary matrix corresponding to all of the feasible PRB allocation patterns and one cost or reward vector based on the selected utility function. The binary matrix, which is named the constraint matrix and shown by A , has M rows regarding the number of available PRBs and $C \times K$ columns corresponding the number of feasible allocation patterns for each user (C) and number of users (K). Each entry of this matrix has a binary value which indicates whether a certain PRB is assigned to the associated user or not. The idea can be described by a simple example. Suppose that there are four PRBs and two users ($M = 4, K = 2$). For a given user, by ignoring the PRB allocation of the other users, in this case there are a few feasible allocation patterns that can be allocated to the given user. Now for the particular user i , the constraint matrix (A_i) will be shown as:

$$\mathbf{A}_i = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad (3.2)$$

The first column in Equ 3.2 shows that no PRB is assigned to user i , the second column states only that the first PRB is assigned to user i , and the last column expresses that all of the available PRB are assigned to user i and there is no remaining PRB for the other user. The number of columns in matrix A_i is 11 in our example ($C = 11$). In general, the total number of

columns for each user is

$$C = 1 + \sum_{m=1}^M (M - (m - 1)) = \frac{1}{2}M^2 + \frac{1}{2}M + 1 \quad (3.3)$$

The system constraint matrix (A) is two replicas of the A_i because there are two users in the system and can be shown as:

$$\mathbf{A} = \left[\begin{array}{c|cccccccccccc|cccccccccccc} & \text{user 1} & & & & & & & & & & & \text{user 2} & & & & & & & & & \\ \hline & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{array} \right] \quad (3.4)$$

It is worthwhile to mention three important points with respect to the constraint matrix (A): (1) each pattern in this matrix contains the contiguity constraint and (2) for each user just one of the patterns from A_m has to be selected as well as (3) each PRB should be allocated to at most one user. The cost or reward vector (R) is calculated based on the selected utility function and scheduling strategy for each column of matrix A . For each column of matrix A , the allocated PRBs for the associated user are determined and the utility function can be calculated according to the known utility function. Therefore, the reward vector has $C \times K$ different elements. These search-based algorithms use *set partitioning* approach to solve the scheduling problem. The reward vector for the given example is shown as follows:

$$\mathbf{R} = \begin{bmatrix} R_{1,1} & \cdots & R_{1,C} & R_{2,1} & \cdots & R_{2,C} \end{bmatrix} \quad (3.5)$$

where in $R_{i,j}$, i denotes user index and j pattern index.

3.4 The Size of Search Space

In this section, the number of feasible solutions to allocate M PRBs among K users is calculated. This calculation just considers the FDPS search space. Two scenarios can be regarded in computation of search space. This section specifies how large the search space of scheduling can be in the allocation problem. In the following parts of this section, it is assumed that there are K active users and M available PRBs. In practice, the set of allowed value of M is given as $\{6, 15, 25, 50, 75, 100\}$ based on Table 2.2 concerning the selected bandwidth.

3.4.1 Scenario 1: assignment of the whole PRBs among users

In this scenario, it is assumed that all of the PRBs are assigned to the users and there is no unallocated PRB after scheduling. At first we select μ users out of K and distribute the M PRBs among these users. Due to the contiguity constraint in uplink LTE, we should share M PRBs into μ ordered set wherein each set has m_i adjacent PRB, which is assigned to user i . Hence, we should calculate the number of different combinations that satisfies $M = m_1 + m_2 + \dots + m_\mu$.

In combination theory, this problem has $\binom{M-1}{\mu-1}$ different solutions [11] and there are μ permutations of K patterns to select μ users out of K total users by considering the sequence.

Therefore, there are $\binom{M-1}{\mu-1} P(K, \mu)$ possible PRB allocations which μ users employ the M PRB. Adding all the allowable numbers of the users, the search space will be

$$\sum_{\mu=1}^K \binom{M-1}{\mu-1} P(K, \mu) = \sum_{\mu=1}^K \binom{K}{\mu} \mu! \binom{M-1}{\mu-1} \quad (3.6)$$

As a practical case, assuming of 25 PRBs ($M = 25$) and 10 active users ($K = 10$), the search space has 5.26×10^{12} possible allocation patterns and the scheduler should traverse among them and choose the most efficient one. Assume that checking for one possible solution takes 1×10^{-9} seconds. The running time of a complete search is about 5.26×10^3 seconds and this duration is much longer than the maximum time of the scheduling which is one ms. Back to the simplified given example with two users and four PRBs, we have eight allocation combinations which are shown in Table 3.2.

PRB allocation No.	Set of assigned PRBs for user 1	Set of assigned PRBs for user 2	Relevant column in A matrix for user 1	Relevant column in A matrix for user 2
1	\emptyset	$\{1,2,3,4\}$	1	22
2	$\{1\}$	$\{2,3,4\}$	2	21
3	$\{1,2\}$	$\{3,4\}$	6	19
4	$\{1,2,3\}$	$\{4\}$	9	16
5	$\{1,2,3,4\}$	\emptyset	11	12
6	$\{2,3,4\}$	$\{1\}$	10	13
7	$\{3,4\}$	$\{1,2\}$	8	17
8	$\{4\}$	$\{1,2,3\}$	5	20

Table 3.2: A sample UE-PRB allocation

3.4.2 Scenario 2: assignment of all or some PRBs among users

In this scenario, it is assumed that either all or some of the PRBs are assigned to the users and it is likely that, after scheduling, some of the PRBs are not allocated to the users. Again we assume μ out of K users are chosen for scheduling. The set of allocated PRBs for each of μ user is shown as a_i where i is the index of the user. We arrange these sets in order of PRB number and define starting the PRB number (a_i^s) and finishing (a_i^f) PRB number for each of the different μ sets. Due to the contiguity constraint in uplink, we have

$$1 \leq a_1^s \leq a_1^f < a_2^s \leq a_2^f \dots < a_\mu^s \leq a_\mu^f \leq M \quad (3.7)$$

as noted M is the number of available PRBs. By a little manipulation of Equation 3.7:

$$1 \leq a_1^s < a_1^f + 1 < a_2^s + 1 < a_2^f + 2 \dots < a_\mu^s + \mu - 1 < a_\mu^f + \mu \leq M + \mu \quad (3.8)$$

Thus, the number of choices of μ sets of contiguous PRBs is equal to the number of 2μ integer that satisfy the $1 \leq b_1 < \dots < b_{2\mu} \leq M + \mu$. By using combination theory, there are $\binom{M + \mu}{2\mu}$ different solutions for this equation. By considering of the distribution of these μ

sets to μ users, there are $\binom{M + \mu}{2\mu} P(K, \mu)$ possible PRB allocations which μ users employ the M PRBs. Adding all the allowable numbers of the users, the search space will be

$$\sum_{\mu=0}^K \binom{M+\mu}{2\mu} P(K, \mu) = \sum_{\mu=0}^K \binom{K}{\mu} \mu! \binom{M+\mu}{2\mu} \quad (3.9)$$

As a practical case, assuming 25 PRBs ($M = 25$) and 15 active users ($K = 15$), the search space is more than 10^{21} and the running time of a complete search is unacceptable compared to the duration time of the scheduler (1ms). Back to the simplified given example with two users and four PRBs, we have 51 different allocation combinations.

3.5 Scheduling Strategy

In this section, different allocation strategies are introduced for LTE systems. The scheduling policy determines the metric formula for matrix-based algorithms and the reward formula for pattern-based algorithms. All of the metric functions can be broken into four main categories: (1)channel-unaware, (2)channel-aware/QoS-unaware, (3)channel-aware/QoS-aware and (4)power-aware. In the following, some of the most common metric functions in each category are introduced.

3.5.1 Channel-unaware

This category of metrics is widely used in wired networks where the media is time-invariant. In wireless networks, this type of metrics has less efficiency than other types due to the time-variation of the channel.

1) First In First Out (FIFO): in this allocation policy, users are served according to the order of resource requests. The corresponding metric of this policy can be expressed as

$$M_{i,m}^{FIFO} = t - T_i \quad (3.10)$$

where t is the current time and T_i is the time when the request was issued by user i .

2) Round Robin (RR): the RR metric is the same as the FIFO metric with the difference that T_i refers to the last time when the user was served. This policy is almost fair in terms of time which is shared among users not user throughput.

3) Blind Equal throughput (BET): the throughput fairness among users can be achieved by using this scheme. The metric of this scheme is

$$M_{i,m}^{BET} = \frac{1}{\overline{R}_i(t-1)} \quad (3.11)$$

where $\overline{R}_i(t)$ is the achieved average throughput until current time t by the user i . $\overline{R}_i(t)$ is calculated by

$$\overline{R}_i(t) = (1 - \frac{1}{T_w})\overline{R}_i(t-1) + \frac{1}{T_w}r_i(t) \quad (3.12)$$

where T_w is the scheduling time window size (usually in the order of 1000), and $r_i(t)$ is the achieved data rate of user i at time t . In this scheme, BET assigns resources to the users that have lower average throughput rather than other users. As is obvious, this policy does not care about the arrival rate of the users and its goal is only to equalize the moving average throughput among users.

4) Weighted Fair Queuing (WFQ): this approach both includes user priority and avoids the possibility of users' starvations. A sample approximation metric of WFQ is expressed as

$$M_{i,m}^{WFQ} = w_i \cdot M_{i,m}^{RR} \quad (3.13)$$

where w_i is the specific weight of user i related to the associated priority of that user and $M_{i,m}^{RR}$ is the RR metric explained before. In other words, the scheduler allocates the resources to the users with higher priority and shorter waiting time.

5) Earliest Deadline First (EDF): this approach is a type of guaranteed delay scheme and its goal is to assign the resources in such a way that all of the packets are received within a certain deadline. To accomplish this goal, the metric has to include both the time when the packet is received and the allowable deadline for the packet. EDF, as its name itself clearly states, allocates first users who have the closest deadline expiration. Mathematically, the EDF metric can be formulated as

$$M_{i,m}^{EDF} = \frac{1}{\tau_i - D_{HOL,i}} \quad (3.14)$$

where τ_i is the delay threshold for the user i and $D_{HOL,i}$ is the head of line delay that means the delay of the first packet to be transmitted by the user i .

6) Largest Weighted Delay First (LWDF): in the delay-aware schemes, all of the packets which expire after the allowable deadline are dropped. This scheme includes the acceptable packet loss rate into the metric as well as the head of line delay and delay threshold of the users. The metric can be calculated as

$$M_{i,m}^{LWDF} = \alpha_i \cdot D_{HOL,i} = -\frac{\log \delta_i}{\tau_i} \cdot D_{HOL,i} \quad (3.15)$$

On the other hand, α_i acts like a weight for the LWDF metric which is calculated by considering both the acceptable packet loss rate and delay threshold.

3.5.2 Channel-aware/QoS-unaware

Thanks to CQI feedback, the scheduler can estimate the channel quality between users and eNB. With knowledge of the channel Signal to Noise Ratio (SNR) between users and eNB, the maximum achievable throughput can be predicted by using either the AMC tables or Shannon channel capacity formula as

$$d_i^m(t) = \log[1 + SNR_i^m(t)] \quad (3.16)$$

where $d_i^m(t)$ is the expected achievable throughput for the user i over the PRB m .

1) Maximum Throughput (MT): the aim of this scheme is to maximize the overall throughput of the system without regard for QoS provisioning and fairness among users. Its metric can be shown as

$$M_{i,m}^{MT} = d_i^m(t) \quad (3.17)$$

In this way, the allocation in the uplink is not as simple as that in downlink due to the contiguity constraint. The uplink scheduler should do a comprehensive search among all of the feasible allocation patterns to come up with a pattern which maximizes the following expression

$$\max \sum_{i=1}^K \sum_{m \in a_i} M_{i,m}^{MT} = \sum_{i=1}^K \sum_{m \in a_i} d_i^m(t) \quad (3.18)$$

where a_i is the set of all of the assigned PRBs to the user i . This scheme just focuses on maximization of cell-throughput and suffers from fairness among users in terms of throughput.

2) Proportional Fair (PF): in general, this approach includes fairness and spectral efficiency simultaneously. Its metric is obtained by combining those of MT and BET as follows

$$M_{i,m}^{PF} = M_{i,m}^{MT} \cdot M_{i,m}^{BET} = d_i^m(t) / \bar{R}_i(t-1) \quad (3.19)$$

in terms of fairness, this scheme is between MT(without fairness) and BET(complete fairness). In this scheme, the parameter T_w in Equation 3.12 plays an important role which determines the window size over which fairness wants to be executed. It is worth pointing the difference between $d_i^m(t)$ and $r_i(t)$, where $d_i^m(t)$ is the expected (predicted) data-rate of user i over PRB m at time t while $r_i(t)$ is the actual achieved data rate of user i at time t . On the other hand, at the particular time t scheduler knows the last achieved data rate for all users (i.e. all of the $r_i(t-1)$) and based on the CQI feedback can predict the expected data rates for current time (i.e $d_i^m(t)$). The Generalized Proportional Fair metric can be developed as an extended version of the PF metric by introducing two new parameters, ξ and ψ

$$M_{i,m}^{GPF} = \frac{[d_i^m(t)]^\xi}{[\bar{R}_i(t-1)]^\psi} \quad (3.20)$$

By changing the values of ξ and ψ , there is an effect on the instantaneous data rate and past achieved data rates on the metric. This metric is an exhaustive metric which covers different scheduling policies such as PF metric ($\xi = \psi = 1$), BET metric ($\xi = 0$) and MT metric ($\psi = 0$). In this developed metric, these two new parameters can be either fixed or adaptive. In the adaptive GPF scheme, ξ and ψ are updated depending on the system condition to tune the achievable fairness level.

3)Throughput to Average (TTA): This approach can be considered an intermediate between MT and PF. Its metric is

$$M_{i,m}^{TTA} = \frac{d_i^m(t)}{d_i(t)} \quad (3.21)$$

where $d_i(t)$ is the expected achievable throughput for the user i over the entire bandwidth. To predict $d_i(t)$, at first the effective SNR of the certain user i should be calculated. In downlink, Exponential Effective SNR Mapping (EESM) and Mutual Information Effective SNR Mapping (MIESM) methods are used to convert SNR values of the PRBs into one effective SNR value in

an additive gaussian White noise channel [12]. For the EESM method [13] the effective SNR of user i is calculated by

$$\gamma_i = \beta_z \text{Ln} \left(\frac{1}{|N_i|} \sum_{m \in N_i} \exp\left(\frac{\gamma_{i,m}}{\beta_z}\right) \right) \quad (3.22)$$

where $\gamma_{i,m}$, N_i and z are the SNR of user i over PRB m , the set of assigned PRBs to user i and the index of selected MCS, respectively. $|\cdot|$ operator returns the size of inside set. β_z is the adjusting factor corresponding to the selected MCS which can be obtained from [14]. For the MIESM method [15, 16], the effective SNR can be obtained by

$$\gamma_i = I_z^{-1} \left(\frac{1}{|N_i|} \sum_{m \in N_i} I_z(\gamma_{i,m}) \right) \quad (3.23)$$

where I_z is the mutual information function which depends on the specific modulation alphabet z and can be computed from [15, 16].

In uplink, the SNR per sub-carrier is not directly related to the data symbol. This is because of the SC-FDMA transmission, which spreads each data symbol over the whole bandwidth (see Figure 2.3). The effective SNR of an SC-FDMA symbol cannot be approximated using EESM or MIESM (as in OFDM), but rather it can be approximated as the averaged SNR over the transmission bandwidth (i.e., the sum of SNR over the different PRBs, divided by the number of PRBs) divided by the average interference over the transmission bandwidth [17] as

$$\gamma_i = \frac{1}{|N_i|} \sum_{m \in N_i} \frac{\gamma_{i,m}}{|N_i|} \quad (3.24)$$

3.5.3 Channel-aware/QoS-aware

By increasing the high rate demands, the need for transmissions with QoS is unavoidable. It is worthwhile to note that QoS-aware does not necessarily mean QoS provisioning. It means the scheduler makes allocation decisions depending on the user quality requirement without necessarily guaranteeing the users' requirements.

1) Guaranteed Data Rate schedulers: the most well-known category of QoS-aware schedulers are guaranteed data rate ones. A general TDPS/FDPS sample of this category is proposed in [18]. In this scheme, the user is divided into two sets: users whose data rates are below the

associated target data rates and users who satisfy the target data rates. Users belonging to the first and second sets are prioritized by using BET and PF metrics. After prioritization of the users, a number of candidate users has been selected for the FDPS phase. FDPS performs PRB allocation based on the PF Scheduled (PFsch) metric as

$$M_{i,m}^{PFsch} = d_i^m(t) / \overline{R_i^{sch}}(t-1) \quad (3.25)$$

where $R_i^{sch}(t-1)$ is similar to Equation 3.12 with the difference that it is updated only when the user i is actually served. Another approach is followed in [19] where the authors prioritized the users at each sub-frame depending on head of line and delay threshold by using the following formula

$$P_i = D_{HOL,i} / \tau_i \quad (3.26)$$

After selecting the user with the highest priority, the scheduler assigns resources to that user to reach the guaranteed bit rate. Then if some resources are left free, the same operation is done for the next user in the priority list. This procedure is continued until all of the resources are allocated.

2) Guaranteed Delay Requirements schedulers: the aim of this category of scheduler is to guarantee the delay requirement for users. As noted, each user has different types of data traffic (flow). In a simple case, two types of flow are considered: real-time flow, which has an associated delay requirement, and non-real-time flow without any bounded delay.

The Modified LWDF (M-LWDF) is a channel aware version of LWDF which was explained before. The metric is the weighted PF where weight is determined by head of line delay for real time flows. In other words, the metric is

$$M_{i,m}^{M-LWDF} = \alpha_i \cdot D_{HOL,i} \cdot M_{i,m}^{PF} = \alpha_i \cdot D_{HOL,i} \cdot \frac{d_i^m(t)}{\overline{R_i}(t-1)} \quad (3.27)$$

M-LWDF metric offers a good balance among spectral efficiency, fairness and QoS provisioning, by using the channel quality information.

Another scheme, which is a combination of the PF metric and delay bounded metric, is presented in [20]. In this scheme, the Exponential/PF metric for real time flows are computed

as

$$M_{i,m}^{EXP/PF} = \exp\left(\frac{\alpha_i \cdot D_{HOL,i} - \chi}{1 + \sqrt{\chi}}\right) \cdot \frac{d_i^m(t)}{\bar{R}_i(t-1)} \quad (3.28)$$

3.5.4 Power-aware

Nowadays, green networking is a hot topic for both researchers and mobile operators. The goal of green networking is to minimize power consumption of network structures to ensure eco-sustainability. Without regarding the ecological effect, power consumption is an important issue in uplink compared to downlink, since the transmitter units of the uplink are UEs with limited energy batteries. In downlink, a simple way to reach this goal is to maximize the spectral efficiency (employing MT metric). With high data rates, a given amount of data can be transmitted during a low time interval that leads to eNb switches more frequently to the sleep mode. To the best of my knowledge, there are a few research studies regarding power-aware schedulers in uplink LTE and almost all of them are pattern based. In the next chapter, a new scheme for this scheduling is presented in detail.

3.6 Literature Review in Uplink LTE Scheduling

In this section, some of the previous works regarding to uplink LTE scheduling are investigated. One of the first works in uplink scheduling is presented in [21]. In this paper the objective is to derive low complex algorithms for channel dependent scheduling to maximize sum data rate in uplink LTE. The algorithms consist of PRB or chunk (a subset of PRBs) assignments and power allocations for multiple chunks with constrained transmit power to the UEs. Authors consider Minimum Mean Square Error (MMSE) equalizer. From [22, 23], in MMSE the effective SNR of each user can be written as

$$\gamma_i = \left(\frac{1}{\frac{1}{|N_i|} \sum_{m \in N_i} \frac{\gamma_{i,m}}{\gamma_{i,m} + 1}} - 1 \right)^{-1} \quad (3.29)$$

The authors show that there is an increase up to 130% in sum rate capacity by using the proposed scheduling algorithm relative to RR scheme. This work is extended to include the impact of imperfect channel information on the scheduling in [24]. These two works suffer from un-

fairness disadvantage. To address this drawback authors use the logarithmic user data rate as a utility function provides proportional fairness as shown in [25].

Calabrese *et. al* in [26] provides a search-tree-based channel- aware packet scheduling algorithm. The allocation is performed by searching and choosing the path, within the tree, with the highest system metric. This algorithm introduces a critical variable named out-degree. We exemplify the main idea of the algorithm and effect of out-degree parameter with a simple case. Assuming three UEs and three PRBs with metric matrix shown in Table 3.3

	PRB_1	PRB_2	PRB_3
UE_1	$M_{1,1} = 380$	$M_{1,2} = 670$	$M_{1,3} = 1530$
UE_2	$M_{2,1} = 300$	$M_{2,2} = 730$	$M_{2,3} = 1390$
UE_3	$M_{3,1} = 650$	$M_{3,2} = 810$	$M_{3,3} = 1280$

Table 3.3: Example of UE-PRB metric matrix

By setting the out-degree (Deg) to one, the algorithm has the following procedure:

1. Find the UE-PRB pair with the highest metric value
2. Assign that PRB to the associated UE
3. Delete the assigned UE and corresponding PRB from the metric matrix
4. Repeat from 1 until all of the PRBs are assigned

By utilizing this straightforward algorithm, PRB1, PRB2 and PRB3 are assigned to UE2, UE3 and UE1 respectively, and the total metric value is 2640 ($1530+810+300=2640$). Next we set out-degree to two and form the associated tree as follows as shown, the total metric is 2910 which is greater than the previous one at the expense of increasing the complexity and computational time. This scheme assigns a fixed size of bandwidth for each user which results in decreasing the system performance. Therefore, the authors in [27] present an adaptive transmission bandwidth based scheduling to cover the previous problem. As a result, in this new scheme the uplink data rate is increased by approximately 20% in average cell throughput compared to a fixed bandwidth channel-aware approach. Calabrese *et. al* in [28] explore the performance of frequency and time domain scheduling in LTE. In particular, they compare various scheduling metrics in terms of average cell throughput and outage user throughput. Lee

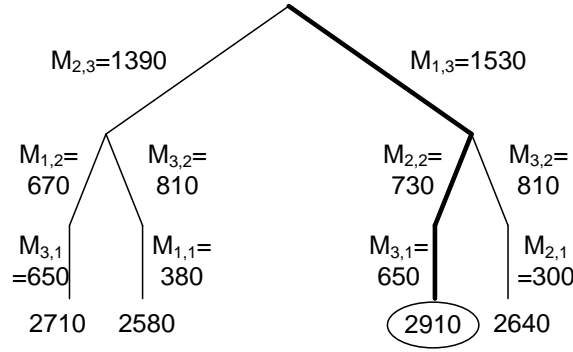


Figure 3.2: Associated tree for given example. Thick line indicates the assignment

et. al in [1] studied the FDPS problem and proposed four different matrix-based algorithms which consider contiguity constraint in uplink LTE. The metric value of these approaches is logarithmic data rate to include fairness and maximize cell throughput. At first, the effect of contiguity constraint on the scheduling is compared. Consider a sample case that is shown in Figure 3.3. In this Figure, each element denoted the PF metric value for the corresponding PRB and user. The most efficient allocations of this case are shown in Figure 3.4 in downlink and

users\PRBs											
A	8	7	6	5	4	3	4	5	6	7	8
B	1	8	1	8	2	8	3	8	2	7	1
C	6	6	6	5	5	6	4	4	6	6	5
D	3	4	5	6	7	8	9	8	7	6	5
E	7	8	6	3	6	4	5	8	2	8	6

Figure 3.3: A sample metric matrix

uplink, respectively. The difference between these two scheduling arises from contiguity constraint which should be considered in uplink. In downlink without contiguity the total metric is 85 while in uplink this value is 83 which is obviously less than 85 due to using SC-FDMA in uplink. In this case, to come up with the best solution in uplink, the scheduler should search among all of 42505 (based on Equation 3.6) feasible pattern allocations, calculate the total metric and select the pattern with the highest metric value as final solution.

Clearly, finding the best solution among all of the possible patterns during 1ms sometimes is impractical. This is why researchers try to find a heuristic algorithm that approximates

users\PRBs	Without contiguity constraint											users\PRBs	With contiguity constraint										
A	8	7	6	5	4	3	4	5	6	7	8	A	8	7	6	5	4	3	4	5	6	7	8
B	1	8	1	8	2	8	3	8	2	7	1	B	1	8	1	8	2	8	3	8	2	7	1
C	6	6	6	5	5	6	4	4	6	6	5	C	6	6	6	5	5	6	4	4	6	6	5
D	3	4	5	6	7	8	9	8	7	6	5	D	3	4	5	6	7	8	9	8	7	6	5
E	7	8	6	3	6	4	5	8	2	8	6	E	7	8	6	3	6	4	5	8	2	8	6

Figure 3.4: Allocation difference between uplink and downlink

optimal solution with lower complexity and accordingly lower computational time. Lee *et. al* introduce four heuristic algorithm and compare them with each other in terms of short-term and long-term fairness as well as cell throughput.

1) *Carrier by carrier in turn*: in this algorithm, scheduler assigns PRBs from the first PRB to the last PRB consecutively. The starting PRB is the rightmost one. For each PRB, at first the scheduler selects the maximum PF metric value and assigns that PRB to the corresponding user if one of these two conditions meets: (a) none PRB is assigned to the corresponding user, and (b) the previous PRB is assigned to the corresponding user. With this procedure, the scheduler assigns all of the PRBs. For the given example, the result of the carrier by carrier algorithm is shown in Figure 3.5. For the first PRB the maximum metric value is 8 and because no PRB has

users\PRBs	8	7	6	5	4	3	4	5	6	7	8
A	8	7	6	5	4	3	4	5	6	7	8
B	1	8	1	8	2	8	3	8	2	7	1
C	6	6	6	5	5	6	4	4	6	6	5
D	3	4	5	6	7	8	9	8	7	6	5
E	7	8	6	3	6	4	5	8	2	8	6

Figure 3.5: Carrier by carrier method

been assigned to the corresponding user before, the scheduler assigns this PRB to user A. The maximum value of the second PRB is 8 and relevant to user B. Like the previous step, because no PRB is assigned to the corresponding user, we assign this PRB to user B. At this time, the scheduler should delete user A, because this user is not the same as the last scheduled user (user B). The scheduler performs this procedure in turn. If before reaching the last PRB, just one user remains, the scheduler assigns all of the remaining PRBs to the last user. Because in

this algorithm, the scheduler does not assign the largest PF metric value at first, and start from one side in sequence, the output may be far from the optimal solution. Figure 3.6 shows one bad example of this algorithm. In this Figure, there are 2 users and 11 PRBs and L is a large

users\PRBs											
A	1	0	L	L	L	L	L	L	L	L	L
B	0	1	0	0	0	0	0	0	0	0	0

Figure 3.6: Bad example of carrier by carrier method

number. As shown, total metric value of this algorithm is 2 but the optimal total metric value is $9 * L + 1$. This means this algorithm was trapped by this example.

2) *Largest-metric-value-PRB-first*: it is shown from Algorithm 1 that scheduling PRBs in sequence from one end side does not provide high efficiency. The drawback of the first algorithm is that it does not consider the largest value at first. The second algorithm has two key ideas: it considers largest value at first and packs large items. The procedure is that at first the scheduler finds the largest value and its corresponding user and then finds the second largest value for that user, and finally assigns all of the PRBs in between to corresponding user. The algorithm is explained by the given example. Figure 3.7 shows the result of scheduling problem by using Algorithm 2. The largest value in this matrix is 9 and the relevant user is

users\PRBs											
A	8	7	6	5	4	3	4	5	6	7	8
B	1	8	1	8	2	8	3	8	2	7	1
C	6	6	6	5	5	6	4	4	6	6	5
D	3	4	5	6	7	8	9	8	7	6	5
E	7	8	6	3	6	4	5	8	2	8	6

Figure 3.7: Largest-metric-value-PRB-first method

D. Now the scheduler finds the next maximum value for user D: this value is 8. Because the corresponding PRBs for first and second largest value are adjacent to each other, the scheduler assigns both of these PRBs to user D and deletes user D for the remaining scheduling process. Now the scheduler finds the maximum value from the remaining matrix (so far the whole row

for user D and 6th and 7th columns are deleted). The maximum value is 8 (intersection of first column and first row) and the next maximum value for user A is again 8 (intersection of last column and first row), but now we cannot assign all the PRBs in between to user A, inasmuch as 6th and 7th PRBs are already assigned to user D. Therefore, the scheduler searches for the next largest value. This algorithm solves the problem presented for the bad example of first algorithm. Figure 3.8 shows a bad example for this algorithm. As shown, the first and last PRB

users\PRBs										
		L+1	0	0	0	0	...	0	0	L+1
		0	L	L	L	L		L	L	0

Figure 3.8: Bad example of largest-metric-value-PRB-first method

have the largest metric value and all of the PRBs are assigned to just user A.

3) *Riding peaks*: in channel-aware scheduling strategies, metric value depends on channel SNR. Also in multi user mobile network, channel SNR values are correlated both in time and frequency. Correlation of SNR values in frequency means that if user i at PRB m has large metric value, that user with high probability has high metric value in PRBs $m - 1$ and $m + 1$ [29]. This algorithm also selects the largest metric value and augments that metric by one neighbour PRB. The resulting outcome for given example is depicted in Figure 3.9. The largest

users\PRBs												
		8	7	6	5	4	3	4	5	6	7	8
		1	8	1	8	2	8	3	8	2	7	1
		6	6	6	5	5	6	4	4	6	6	5
		3	4	5	6	7	8	9	8	7	6	5
		7	8	6	3	6	4	5	8	2	8	6

Figure 3.9: Riding peaks method

value in this matrix is 9 and the relevant user is D. Because until now no PRB is assigned to user D, the scheduler allocates this PRB to user D. The second largest value is 8 (intersection of first column and first row) and the corresponding user of this element is user A. Inasmuch as no PRBs is assigned to user A so far, we assign the first RB to user A. The next largest

value is in intersection of the last column and the first row. For this element, one PRB is assigned to the corresponding user, and this assigned PRB is not adjacent to the relevant PRB of the selected element. So the scheduler cannot assign this PRB (last PRB) to user A. The scheduler performs likewise until all of PRBs are assigned. This algorithm is named riding peak because at first high value PRBs are assigned to the users and later the remaining PRBs are allocated. This algorithm attains the optimal result for the bad example of algorithm 2. There are still bad examples that trap this algorithm and the results are completely far from optimal assignments. Figure 3.10 shows a bad example for algorithm 3. In this example, at

users\PRBs											
A	L+1	0	L	L	L	L	L	L	L	L	L
B	0	L+1	0	0	0	0	0	0	0	0	0

Figure 3.10: Bad example of riding peaks method

first the peaks are assigned to the users and then the remaining PRBs are allocated such that the contiguity limitation would not be violated.

4) *PRB grouping*: in LTE, channel qualities are correlated both in time domain and in frequency domain. Unfortunately the strength of correlation in frequency domain is weaker than time domain [30]. In other words, SNR values are correlated in frequency domain generally but the granularity of correlation is not as small as one PRB. This means that sudden changes in metric value may leads to output efficiency becomes low. This situation is shown in Figure 3.11. In this Figure, one instantaneous peak results in bad assignment of user B. Also one instantaneous drop of user A limits the PRB allocation of this user. To overcome this problem, one approach is to extend one RPB to a group of PRBs and apply algorithm 3 to the grouped PRBs. In the example of Figure 3.11 (or Figure 3.10), if the scheduler considers a group of x contiguous PRBs (e.g. $x = 3$) instead of one PRB, then the scheduler has a view wide enough to obtain an optimal solution. Thus, this PRB grouping seems to solve the previous problem. On the other hand, algorithm 4 is a version of algorithm 3 in which at first we grouped PRBs and then applies algorithm 3. The number of PRBs (x) that are grouped with each other depends on both number of users and number of PRBs available. The resulting outcome for the given example is depicted in Figure 3.12. In this example there are 11 PRBs and 5 users and

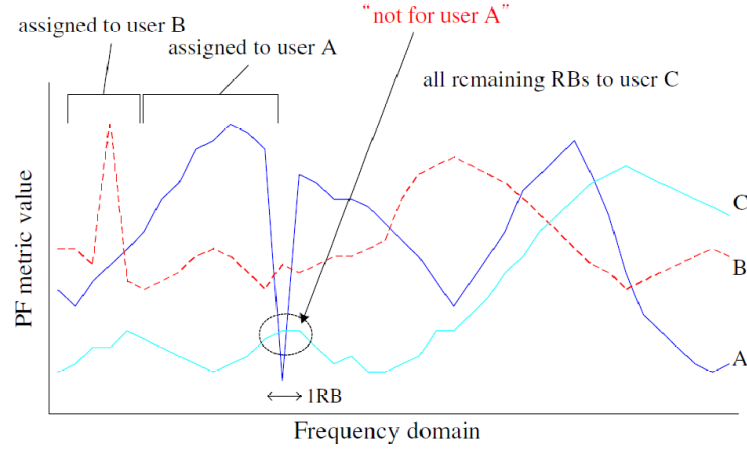


Figure 3.11: The drawback of riding peaks method [1]

users\PRBs												
A	8	7	6	5	4	3	4	5	6	7	8	
B	1	8	1	8	2	8	3	8	2	7	1	
C	6	6	6	5	5	6	4	4	6	6	5	
D	3	4	5	6	7	8	9	8	7	6	5	
E	7	8	6	3	6	4	5	8	2	8	6	

Figure 3.12: PRB grouping method

assume that $x = 2$. At first, the scheduler groups two adjacent PRBs with each other such that the aggregate PF metric value of each group is equal to summation of two adjacent PRBs, and then applies algorithm 3. There are some examples that can cheat algorithm 4, but these situations do not happen in practice. After a comprehensive simulation, the authors conclude that PRB-grouping algorithm gives us the closest outputs to optimal solution and has the best consequences in terms of sum throughput and fairness.

In [31] two heuristic algorithms are introduced. The main idea of these two approaches is to assign a PRB to the user with maximum marginal utility value, where the marginal utility represents the gain in the utility function when a selected PRB, m , is allocated to specified user, i , compared to the utility of user i before the allocation of PRB m . Another work is carried out in [32] in resource allocation to improve the sum throughput of the users. In this work, the authors consider practical scenarios and assign appropriate modulation and coding schemes to the users

based on the channel quality and moreover bandwidth sharing. In [33], Kim *et. al* proposed an algorithm in which the scheduler chooses a chunk of PRBs with the highest channel gain difference between best user and second best user after equalizing the number of users and the number of chunk PRBs. In this study, the metric value is throughput by using Shannon Theory. To evaluate the performance of the proposed algorithm, the authors compare their approach with the carrier-by-carrier and PRB-grouping methods which are described before and show that the proposed scheme is well-suited for throughput maximization of SC-FDMA. In [34], the researchers propose a QoS uplink scheduling algorithm for LTE collaborating with delay estimation by using Equation 3.27 as metric function.

Delgado *et. al* expressed two highly scalable heuristic algorithms with maximum delay and minimum required throughput constraints in [35]. Their performance was analyzed not only in terms of throughput, resource allocation and fairness, but also in terms of delay and number of users effectively served. Another delay bounded scheduling is presented by Li *et. al* in [36]. In this study, the objective was to minimize power transmission for uplink systems of LTE with constraints on mean queuing delay. The scheduler takes into account CSI and QSI in the allocation scheme. All of the aforementioned heuristic algorithms were matrix-based. In [37], Wong *et. al* present a novel reformulation of the scheduling problem as a pure Binary Integer Program (BIP) called the set partitioning problem. In this thesis, this scheme is named pattern-based approach. The main idea is that the scheduler makes scheduling decisions based on a reward function. In other words, each feasible allocation scheme maps to a reward value and finally the most efficient pattern, which has minimum or maximum reward value, is selected. Wong *et. al* also present a greedy heuristic algorithm that approaches the optimal performance. The objective function is the maximization of cell capacity with constraint on power consumption. This work is extended by Sokmen *et. al* in [38].

In [39], the authors studied sum-power minimization based resource allocation in uplink LTE by utilizing the BIP method. The exponentially complex BIP problem is transformed into a canonical dual problem in the continuous space, which is a concave maximization problem. Based on the solution of the continuous dual problem, an iterative algorithm is proposed that minimizes the sum power by performing joint power and sub-channel allocation while satisfying the users target data rates. Another power-aware work was performed by Dan *et. al* in [40].

The authors proposed a scheduling framework by considering HARQ constraints. Because of the high complexity of the BIP framework, a heuristic power efficient scheduling scheme with a tunable complexity parameter, which trades-off between complexity and efficiency, is proposed.

Chapter 4

Heterogeneous Delay-Power Resource Allocation in Uplink LTE

The advantages of OFDM, such as high spectrum efficiency, robustness to time-dispersive radio channels, and the low complexity of receivers, have made it a good candidate technology for broadband air interface of downlink LTE. This technique suffers from a significant disadvantage where instantaneous transmitted power varies noticeably resulting in large PAPR. In uplink where UEs have limited battery life, this factor plays an important role in resource allocation schemes. To address this drawback, 3GPP adopted the single carrier scheme in the uplink system. In this scheme, multiple sub-channels can be assigned to a certain user if they are contiguous to each other. This scheme leads to lower power consumption in UEs, and longer battery life [41]. In recent years, with the growth of mobile Internet, new mobile applications have been introduced to mobile users. Nowadays, consumers expect high speed data services such as VOIP, online gaming, video conferencing, multimedia streaming, and many others. Delays have crucial effects on the performance of these new high bandwidth demanding applications.

The main goal of this work is to minimize packet delay and power consumption simultaneously. Unfortunately, these two criteria conflict with one another, i.e. by transmitting more packets, the packet delay decreases while transmitting power increases. The focus of this chapter is to devise an approach where packet delays for different classes of data and power consumption of UEs are optimized.

4.1 Queueing Theory Basics

Figure 4.1 shows a simplified model of queue and transmission in wireless systems. This system consists of a finite-length buffer for each user. It is assumed each user has only one flow (data type). The packets enter the buffer at an average rate of $a(t)$ packets per second. If the buffer is full, the packets are dropped and P_{drop} models the long-term average probability of packet dropping. The queue holds up to L packets. The number of packets in the queue at anytime t is $q(t)$. Based on the scheduling strategy, CSI and QSI, the service data rate T is transmitted over the channel. The transmitted packets are subject to errors. The probability of receiving erroneous packets is expressed by P_{loss} .

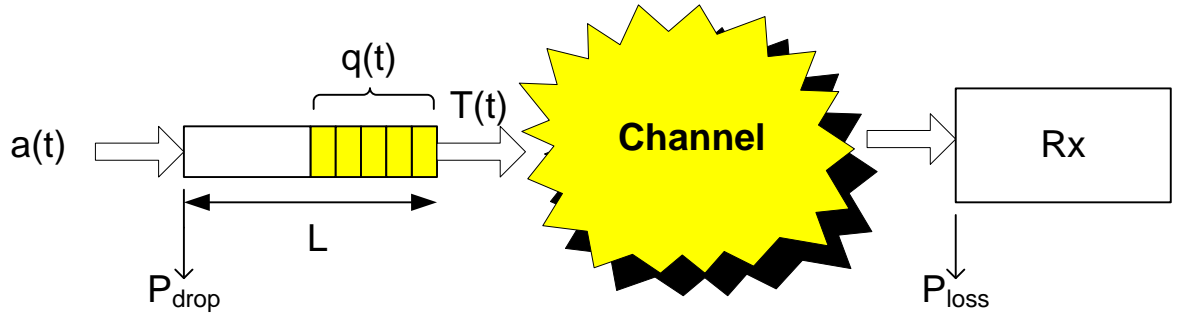


Figure 4.1: Queue model

A queueing system is often described by the notation of A/S/s/k. A stands for the arrival process, such as Poisson, geometric, and deterministic, and S stands for the service distribution, such as exponential, geometric, and deterministic. s denotes the number of servers and k stands for the buffer size where $k = \infty$ when k is absent. In addition, full characterization of the queueing system behaviour requires a description of the service discipline. One of the most well-known arrival models is the Poisson model. In probability theory, a Poisson process is a stochastic process which counts the number of events and the time that these events occur in a given time interval. The time between each pair of consecutive events has an exponential distribution with parameter λ and each of these inter-arrival times is assumed to be independent of other inter-arrival times. Generally, in the Poisson process

$$P\{N(t_1, t_2) = k\} = \frac{e^{-\lambda t} (\lambda t)^k}{k!} \quad (4.1)$$

where $N(t_1, t_2)$ denotes the number of arrivals in an interval (t_1, t_2) and $t = t_2 - t_1$. λ is the expected number of arrivals that occur per unit time.

4.1.1 Little's Law

In queueing theory, Little's law states that the average queue size is equal to the average arrival rate multiplied by the average packet delay [42] as follows

$$\bar{d}_i = \frac{\bar{q}_i}{\bar{a}_i} \quad (4.2)$$

where \bar{d}_i , \bar{q}_i and \bar{a}_i are the average packet delay, average queue size and average arrival rate, respectively and i is the index of the user. This statement is quite general in that it is valid for any probability distributions on arrivals and services as long as the system operates in a first-in-first-out manner.

4.2 System Model

In this chapter, two tasks are carried out by allocation framework: (1) *packet assignment*, i.e. determination of the number of transmitted bits for each user and (2) *PRB assignment*, i.e assignment of PRBs to active users. These decisions are made based on channel quality, desirable Block Error Rates, queue states of buffers and restrictions on the power consumption of users. There are K users in a single cell which communicate with one eNB. In this chapter, the inter-cell interference is neglected and the cell spectrum is divided into M PRBs consisting of 12 consecutive sub-carriers with a 180 kHz bandwidth. All of the scheduling decisions are made in the eNB in every sub-frame. In other words, the basic unit of scheduling is one sub-frame (1 ms) in the time domain and one PRB (180 kHz) in the frequency domain. Scheduling strategy is not specified in LTE, and many researchers have proposed different algorithms according to selected scheduling policies. Most research studies emphasize four policy metrics: reducing transmitted power, increasing aggregate rate, having fairness between users, and minimizing packet delays. Most previous works consider just one type of traffic (flow), but nowadays, with increasing demands for multimedia in cell phones, a more advanced model is needed. In this

chapter, three different traffic types of data (flows) including voice, video, and best effort data, which have their own specifications, are investigated [43].

- *Best effort* data: this traffic type includes applications, such as e-mail and web surfing. This type of data do not impose any requirements on delay and rate.
- *Video* data: this traffic type includes applications, such as video gaming and TV streaming. This class needs guaranteed rate and latency. The arriving data rate of this class is extremely high.
- *Voice* data: this traffic type includes applications, such as voice and Voice Over IP (VOIP). This class needs a guaranteed rate and latency. The arriving data rate of this class is not high. The guaranteed rate of this class is lower than the video class.

Figure 4.2 illustrates the proposed system model of this chapter. The used notations are shown in Table 4.1. In some works, packet assignment and PRB assignment are accomplished separately. These approaches decrease system performance because there is no cooperation between the two parts.

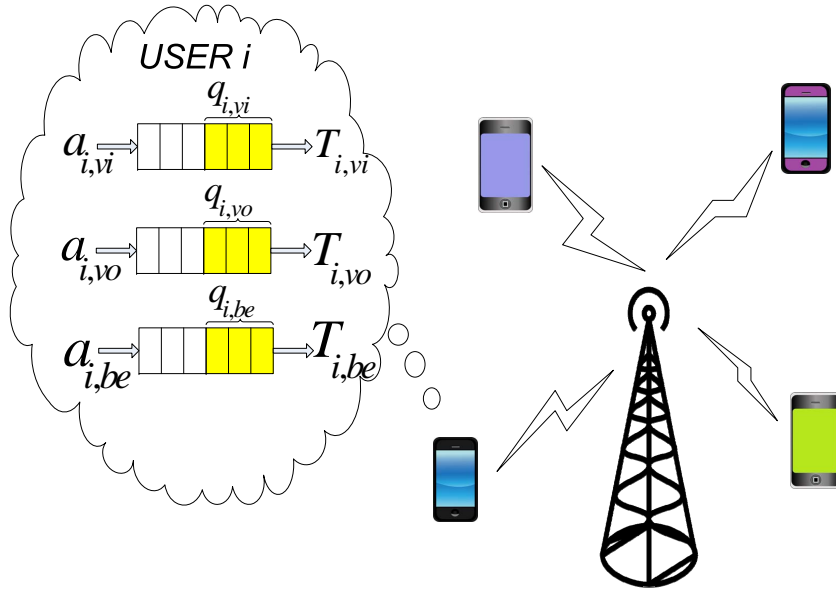


Figure 4.2: System schematic

Parameter	Meaning
K, i	No. of user, user index
M, m	No. of RBs, RB index
j	Traffic type (flow) index
n	sub-frame index
$a_{i,j}$	Arrival packet rate of user i into buffer of flow j
$T_{i,j}$	Service data rate of user i and flow j
L	Buffer length
$q_{i,j}$	Queue size (No. of available bits in queue) of user i and flow j
SDU	Service data unit size
H	No. of header bits for packets

Table 4.1: Summary of notations

Our scheme is categorized as a discrete rate allocation case in which each traffic type of any user can take discrete values from the following set.

$$T_{i,j} \in [0, SDU + H, 2SDU + H, \dots, ySDU + H] \quad (4.3)$$

where y is determined with the value of L , SDU (Service Data Unit) and H . Other notations are given in Table 4.1. As previously noted, the aim is to achieve the best compromise that minimizes both power consumptions and packet delay.

4.2.1 Power criteria

In uplink, because of the limited battery life of UEs, power consumption plays a crucial role in resource allocation problems. To get a higher data rate and lower packet delay transmission, the scheduler should increase transmitted power, which will result in a shortened UE battery life. It is assumed that the SNR of the channel between each UE and eNB and from PRB to PRB is independent. Another important issue related to power is the Block Error Rate (BLER) of transmitting data. There is no closed form formula to model BLER of coded data and so, in this chapter, the measure of Information Outage Probability (IOP) is used as BLER. The

equation is derived from [42]

$$BLER = Q\left(\frac{\log(1 + \bar{\gamma}_i) - \frac{\log_{10} 2 * T_i}{132|N_i|}}{\sqrt{\frac{2\bar{\gamma}_i}{132|N_i|(1+\bar{\gamma}_i)}}}\right) \quad (4.4)$$

where $Q(.)$ is the well known Q function. N_i is the set of PRBs assigned to user i , $|.|$ is the size of the set or the number of PRBs assigned to user i , T_i is the transport block size of user i for all of the traffic types and $\bar{\gamma}_i$ is the required effective SNR of channel for user i to achieve the desired BLER. It is assumed that the desired BLER is known and fixed. T_i is calculated from the following equation in each sub-frame

$$T_i = \sum_{j=1}^3 T_{i,j} = T_{i,video} + T_{i,voice} + T_{i,besteffort} \quad (4.5)$$

The least square approximation method can be used to approximate 4.4 to a simpler version. This algorithm is similar to the approach used by [44]. For BLER=10% it is converted into

$$\bar{\gamma}_i \approx a_x \exp(b_x T_i) - \mu_{0,x} \quad (4.6)$$

where $x = |N_i|$ denotes the number of PRBs used by user i . The values of a_x , b_x and $\mu_{0,x}$ are fixed and given in Table 4.2 for up to 24 PRBs [45]. According to 4.6, the desired effective SNR is just a function of transport block size, and the number of used PRBs for a certain user to achieve the specific BLER.

For each transmission, the measured effective SNR of each user is related to the SNR of the sub-channels, which are assigned to the same user at that sub-frame. Unfortunately, in uplink LTE, because SC-FDMA is being used, we cannot utilize traditional methods, such as EESM and MIESM, to approximate measured effective SNR [17], but it can be computed as average SNR over the assigned sub-channels to the specific user at a particular sub-frame as follows

$$\gamma_i = \frac{1}{|N_i|} \sum_{m \in N_i} \frac{\gamma_i^m}{|N_i|} \quad (4.7)$$

where γ_i^m denotes instantaneous SNR of resource block m seen from user i . Consequently, the

PRBs (x)	a_x	$b_x \cdot 10^3$	$\gamma_{0,x}$
1	1.1748	5.2471	1.1019
2	1.1208	2.624	1.0723
3	1.0977	1.7495	1.0591
4	1.0841	1.3122	1.0512
5	1.0749	1.0498	1.0458
6	1.0682	0.8749	1.0418
7	1.063	0.7499	1.0387
8	1.0588	0.6562	1.0362
9	1.0553	0.5833	1.0342
10	1.0524	0.525	1.0324
11	1.0499	0.4772	1.0309
12	1.0478	0.4375	1.0296
13	1.0459	0.4038	1.0284
14	1.0441	0.375	1.0274
15	1.0426	0.35	1.0265
16	1.0412	0.3281	1.0256
17	1.04	0.3088	1.0249
18	1.0388	0.2917	1.0242
19	1.0378	0.2763	1.0235
20	1.0368	0.2625	1.0229
21	1.0359	0.25	1.0224
22	1.0351	0.2386	1.0219
23	1.0343	0.2283	1.0214
24	1.0336	0.2188	1.0209

Table 4.2: Least-Squares Approximate Model Parameters for BLER=10%

required power to meet the desired BLER is computed from the following equation

$$P_i = k \frac{\text{desired effective SNR}}{\text{measured effective SNR}} = k \frac{\bar{\gamma}_i}{\gamma_i} \quad (4.8)$$

where the nominator is calculated by either 4.4 accurately or 4.6 approximately and the denominator is computed by 4.5 [46]. k is a constant factor to adapt the units of both sides of Equation 4.8.

4.2.2 Packet delay criteria

Over the last decade, with the development of technology, the need for high data rate transmission with low latency and good quality has risen. Thus QSI, packet delay and buffer length

should be considered in scheduling problem. In application, data can be classified into several groups. Each group has its own reasonable arriving rate and maximum tolerable delay. As noted, the average packet delay is a function of average queue size and average arrival rate via the Little theorem as follows

$$\bar{d}_{i,j} = \frac{\bar{q}_{i,j}}{\bar{a}_{i,j}} \quad (4.9)$$

where $\bar{d}_{i,j}$ is the average packet delay for user i for data class (flow) of j , $\bar{q}_{i,j}$ and $\bar{a}_{i,j}$ denote average queue size and average arrival rate corresponding to user i and data class j , respectively. The queue update for each sub-frame is given by

$$q_{i,j}(n) = \min\{q_{i,j}(n-1) + a_{i,j}(n-1) - T_{i,j}(n-1), L\} \quad (4.10)$$

where n and L denote sub-frame index and buffer length, respectively. In practice, the sliding window average with length of W is used instead of taking the average queue size over all of the sub-frames. The sliding window average of length W of variable x at sub-frame $n \geq W$ is calculated by the below formula [47].

$$SWA(x[n], W) = \frac{1}{W} \sum_{l=n-W+1}^n x[l] \quad (4.11)$$

4.3 Problem Formulation

In this section, the resource allocation problem is converted into an optimization problem, then the optimization methods are utilized to achieve the optimal solution. The objective is to optimize packet assignment and PRB assignment to attain the minimum average packet delay and total transmitted power. These two goals conflict with one another since transmitting more data entails a higher power consumption and lower packet delay.

4.3.1 Objective functions and variables

The optimization problem is a function of two criteria: packet delays and transmitted power. Packet delay consists of three various classes of data. Power consumption is formulated as

follows

$$P_t = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{n=0}^t \sum_{i=1}^K \sum_{m=0}^M s_i^m P_i \quad (4.12)$$

where s_i^m is a binary indicator, which denotes a given PRB m is assigned to given user i or not. If PRB m is allocated to user i , this variable is "1", otherwise is "0". P_i is given in 4.8. It is worth stressing that in 4.12, transmitted power (P_t) is a function of four parameters: number of PRBs used by user i ($|N_i|$), required Block Error Rate, transport block size of user i (T_i), and instantaneous SNR of sub-channel m between user i and base station (γ_i^m). Packet delay is formulated as

$$PD = \sum_{i=1}^K \sum_{j=1}^3 \bar{d}_{i,j} = \overbrace{\sum_{i=1}^K \bar{d}_{i,1}}^{\text{video}} + \overbrace{\sum_{i=1}^K \bar{d}_{i,2}}^{\text{voice}} + \overbrace{\sum_{i=1}^K \bar{d}_{i,3}}^{\text{best effort}} \quad (4.13)$$

where j denotes traffic type ($j=1$: video, $j=2$: voice and $j=3$: best effort) and $\bar{d}_{i,j}$ is given in 4.9. Therefore, the target is to minimize 4.12 and 4.13 equations at the same time. In general, the solution is to find a set of transport block size for each user and all sub-frames $\{T_1(n), T_2(n), \dots, T_K(n)\}$ (*packet assignment*), and a matrix of assignment s_i^m for all of the users and PRBs (*PRB assignment*).

4.3.2 Constraints

Allocation constraint: this constraint means that each PRB can be assigned to at most one user in each subframe.

$$\sum_{i=1}^K s_i^m(n) \leq 1 \quad s_i^m \in \{0, 1\} \quad \forall m, n \quad (4.14)$$

Contiguity constraint: SC-FDMA imposes this constraint on scheduling in Uplink LTE. It implies that multiple PRBs can be assigned to one user if these PRBs are adjacent to one another. From [48] the formulation of this constraint can be written as

$$s_i^m(n) - s_i^{m+1}(n) + s_i^x(n) \leq 1 \quad x = m + 2, \dots, M \quad \forall i, m, n \quad (4.15)$$

Queue stability constraint: different definitions of queue stability are available. One necessary condition which should be met to get queue stability is shown in [47]

$$\bar{a}_{i,j} \leq \bar{T}_{i,j} \quad (4.16)$$

Power constraint: based on the LTE standard, the maximum transmission power threshold for each user is 23dBm [41], and therefore

$$\sum_{m=1}^M s_i^m P_i \leq 200 \text{ mW} \quad (4.17)$$

4.4 Resource Allocation Solution

It is not feasible to solve this allocation problem for all sub-frames n in advance. In this context, the approach is to solve the problem in each subframe n separately. Thus, we can drop the index n for notational brevity. As explained in Section 4.3, the scheduling algorithm in this context is not an optimization problem with a single objective function. There are four different objective functions: one for transmitted power given by 4.12 and three for packet delays of the different classes of data. Since each data traffic has its own specification, the packet delay is divided into three objective functions for each class. Therefore, we deal with one multi-objective function problem in this study. The optimal decisions need to be taken by identifying the best trade-offs among these four criteria, which is the goal of the Multi-objective Optimization Problem (MOP). There are several methods to solve MOP. One of the most widely used approaches in practice is the Weighted Sum (WS) method [49]. The idea of the WS method is to convert MOP to a single objective optimization problem. Hence, the WS method yields the following scalar optimization problem for the proposed algorithm

$$\min (f = \alpha_1 w_1 d_{vi} + \alpha_2 w_2 d_{vo} + \alpha_3 w_3 d_{be} + w_4 P) \quad (4.18)$$

where P is P_t for current subframe and is calculated by

$$P = \sum_{i=1}^K \sum_{m=0}^M s_i^m P_i \quad (4.19)$$

d_{vi} , d_{vo} and d_{be} are three separate parts of Equation 4.13. α_j states the priority of delay for data type j . w_z denotes the weight of each individual objective function based on the preference of the scheduler. However, a question may arise as to how to determine the weights. In this work, the idea of specifying w_1 , w_2 , and w_3 is straightforward. The weights are defined as

$$w_{i,j} = \frac{\text{queue size of user } i \text{ and class } j}{\text{buffer length of queue}} = \frac{q_{i,j}}{L} \quad (4.20)$$

It is assumed that for all of the users and classes, the buffer length is fixed L . Obviously, w_j , $j = \{1, 2, 3\}$ takes a value between 0 and 1.0. Two other important points which should be noted regarding w_j , $j = \{1, 2, 3\}$ are: (1) these weights are updated in every sub-frame, and (2) these weights vary for different users. To have the same preference between power and all types of delay, the fixed value of three is selected for w_4 ($w_4 = 3$). The better approach is to evaluate this factor according to the fraction of battery life time of UE multiplied by three. To make the optimization problem easier to solve, we relax queue stability and power constraints. Here, we deal with with two types of decision variables: s_i^m which takes a binary value and T_i which takes a limited discrete integer value.

Wong et al. [37] exploit a *set partitioning* algorithm in uplink SC-FDMA. Wong's work just solves the PRB assignment with the goal of maximizing the total rate of the cell. We adapt the scheme from Wong's work to our proposed algorithm. In general, the problem can be formulated as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{1}_M, \mathbf{A}_{eq}\mathbf{x} = \mathbf{1}_K \end{aligned} \quad (4.21)$$

In this notation, \mathbf{x} is an allocation vector which involves all feasible solutions, \mathbf{A} is a binary inequality (constraint) matrix, \mathbf{A}_{eq} is the binary equality matrix, and \mathbf{c} is the cost vector. Cost vector elements are extracted from the aggregate objective function in 4.18 for each feasible allocation. On the other hand, the algorithm for each user calculates cost function for each feasible allocation pattern and selects the best pattern which has a minimum value among all

the patterns. Matrix \mathbf{A} is given as $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_K]$, where \mathbf{A}_i denotes constraint sub-matrix for user i . With one small example, we describe how to form \mathbf{A}_i . Assume in a sample case that we have three PRBs ($M=3$) and Maximum Transport Block Size (MTBS) for each user for all of the three data classes is two SDU. At first, we analyze the different combinations where we can assign two SDUs among three traffic types. For our example, the combination set is: $\mathcal{P} = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 0), (1, 0, 1), (0, 1, 1), (2, 0, 0), (0, 2, 0), (0, 0, 2)\}$, where the first, second and third part of each element of this set (tuple) denote numbers of SDU for video, voice and best effort traffic type, respectively. We name the size of this set Number of Rate Combination (NRC). In general NRC is calculated by

$$NRC = \begin{pmatrix} \text{MTBS} + 3 \\ \text{MTBS} \end{pmatrix} \quad (4.22)$$

and in every subframe for each user, one of these tuples should be selected. In other words, $\mathbf{A}_i = [\mathbf{A}_i^1, \dots, \mathbf{A}_i^{NRC}]$. Each tuple has several feasible PRB assignment patterns. In our sample, it is

$$\mathbf{A}_i^p = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad \forall i \in \{1, \dots, K\}, \quad \forall p \in \mathcal{P} \quad (4.23)$$

In this matrix, each row represents one PRB and each column corresponds to a feasible PRB assignment pattern. The total number of columns in matrix \mathbf{A}_i^p is

$$C = 1 + \sum_{t=1}^M (M - (t - 1)) = \frac{1}{2}M^2 + \frac{1}{2}M + 1 \quad (4.24)$$

Consequently, matrix \mathbf{A} has M rows corresponding to the number of PRBs and $K \times NRC \times C$ columns. Each column of matrix \mathbf{A} is associated with one binary indicator variable $x_{i,p,c} \in \{0, 1\}$ that denotes whether, for user i , the pattern corresponding to the relevant column is chosen or not. Thus, the allocation vector is defined as: $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_K]^T$; $\mathbf{x}_i = [\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,NRC}]^T$; $\mathbf{x}_{i,p} = [x_{i,p,1}, \dots, x_{i,p,C}]^T$. Also, each column in matrix \mathbf{A} is associated with the cost variable in the cost vector. The value of this variable is the value of the aggregate objective function

given in 4.18. The first constraint term in 4.21 ensures allocation and contiguity constraints, which implies that each PRB can be allocated to just one user and the second constraint term is derived from this rule in which one and only one possible pattern can be selected from \mathbf{A}_i for user i . The proposed optimization problem can be simply solved with any optimization software that has a binary optimization tool.

The formation of allocation vector (\mathbf{x}), cost vector (\mathbf{c}), constraint matrix (\mathbf{A}), equality matrix (\mathbf{A}_{eq}), $\mathbf{1}_M$ and $\mathbf{1}_K$ are explained with a simple example. It is assumed there are two users and three PRBs in the cell as well as MTBS is equal to two. In this example $C = 7$ based on Equation 4.23 and $NRC = 10$ according to Equation 4.22. At first, we form the constraint matrix. This matrix has three rows corresponding to three available PRBs and $2 \times 7 \times 10 = 140$ columns. The constraint matrix (\mathbf{A}) is a combination of two (number of users) sub-matrices (\mathbf{A}_i): $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2]$, where \mathbf{A}_i is a combination of 10 sub-matrices related to a tuple in set \mathcal{P} . In other words, each \mathbf{A}_i^p is related to one of the different combinations that distribute two SDUs to the three various flows of a specified user: $\mathbf{A}_i = [\mathbf{A}_i^1, \dots, \mathbf{A}_i^{10}]$. Now each \mathbf{A}_i^p matrix is shown in 4.23 corresponding to the different PRB allocation patterns for each user. It is worth noting that for each user just one column should be selected. Each column of the constraint matrix maps to a binary variable ($x_{i,p,c}$) where i , p and c denote index of user, distribution of SDUs among three flows and selected PRB allocation. For example if $x_{1,5,6}$ is equal to "1", this means for the first user, the fifth tuple of \mathcal{P} , which assigns one SDU to video and voice flow, is selected. Moreover it determines that the second and third PRBs should be assigned to this user. In this sample case, allocation vector (\mathbf{x}) has $2 \times 7 \times 10 = 140$ (generally $K \times C \times NRC$) rows and one column. $\mathbf{1}_M$ has three (generally M) rows and one column

$$\mathbf{1}_M = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (4.25)$$

The first constraint in 4.24 specifies that each PRB has to be allocated to at most one of the columns in the constraint matrix. This means each PRB can be assigned to at most one user (allocation constraint).

The Equality matrix (\mathbf{A}_{eq}) has two (K) rows and 140 ($K \times C \times NRC$) columns as follows:

$$\mathbf{A}_{eq} = \begin{bmatrix} \overbrace{1 \ 1 \ \dots \ 1 \ 1}^{\text{user1(70 columns)}} & \overbrace{0 \ 0 \ \dots \ 0 \ 0}^{\text{users2(70 columns)}} \\ 0 \ 0 \ \dots \ 0 \ 0 & 1 \ 1 \ \dots \ 1 \ 1 \end{bmatrix} \quad (4.26)$$

Additionally, the $\mathbf{1}_K$ matrix has two (K) rows and one column while the value of all of the elements is "1" as follows

$$\mathbf{1}_K = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (4.27)$$

In summary, the second constraint of 4.21 shows that, for each user, one of the allocation patterns can be selected. The last step is to form the cost vector and each element of this vector is calculated by Equation 4.18 for each allocation pattern of the constraint matrix. For example, the ninth element of this vector, which is related to the ninth column of matrix \mathbf{A} , is calculated as follows: the ninth column of matrix \mathbf{A} refers to the second tuple of set \mathcal{P} , which means the scheduler wants to send only one SDU for best effort flow of user 1. The relevant PRB allocation pattern for this user is to assign the first PRB to this user and the number of used PRB is one. By using Equation 4.6, the desired effective SNR can be computed. Also with knowledge of the CSI and utilizing Equation 4.7, the measured effective SNR is calculated. By employing Equation 4.8, the power part (P) of Equation 4.19 is achieved. The delays are also calculated by Little's law. Now the resulting value of f in Equation 4.19 is the corresponding value of the cost vector.

4.5 Simulation and Numerical Results

To evaluate the proposed framework, we have implemented the proposed algorithm. In our simulation, the total bandwidth consists of 12 PRBs. The channel is viewed as a Rayleigh fading model, and the instantaneous channel SNR is modelled by Exponential distribution with mean SNR equal to 10dB $p(\gamma) = \frac{1}{\gamma_0} \exp(-\frac{\gamma}{\gamma_0})$ and each sub-frame is considered 1ms. The arriving data model is assumed to follow Poisson distribution with an average rate 2, 2, and 1 SDU for video, best effort, and voice data types, respectively. SDU size, buffer length (L), and

average window size (W) are configured to 64 bits, 100 SDU, and four respectively.

Figure 4.3 shows the probability density function of delay for different data types, where ViPD, VoPD and BePD denote video, voice, and best effort packet delays, respectively. The results are extracted from 1000 sub-frames. In this simulation, we assumed $\alpha_1 = \alpha_3 = 0.66$, $\alpha_2 = 1$ and there is one user in the cell. It is worth noting that if MTBS is selected less than the average arrival packets rate (here 5), the packet delays will not converge and queues will not be stable. Figure 4.4 shows the probability density function of delay for different data types while

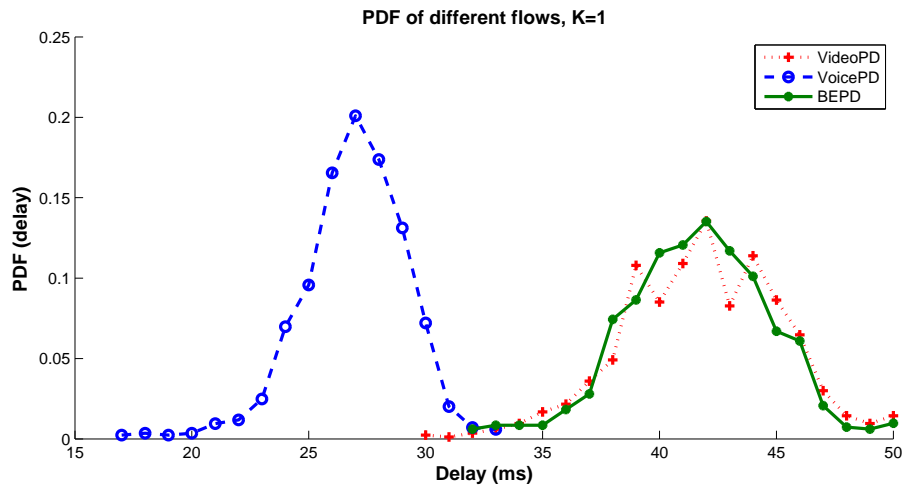


Figure 4.3: PDF of packet delay - one user in the cell

there are two users in the cell. As we expect, the trend is the same as the previous Figure. In Figure 4.5, we compare the average packet delays versus Maximum Transport Block Size (MTBS) for one and two users in the cell. Clearly, as MTBS increases, packet delays decrease. The impact of MTBS on average power per user is shown in Figure 4.6 for one and two users. Hence, having higher MTBS results in better performance. However, by choosing the high value of MTBS, the size of the combination set \mathcal{P} (NRC) was augmented considerably. NRC is proportional to the computation complexity. The effect of MTBS on NRC as a measure of complexity is presented in Figure 4.7. These last five plots illustrate the trade-off between performance and complexity of the scheduler in terms of MTBS parameter. Figure 4.8 demonstrates the effect of varying the number of users on the delays and power consumptions. By increasing the number of active users, the performance deteriorates. It is expected that by increasing the No. of users, the average power per user increases. In Figure 4.8, this is correct in

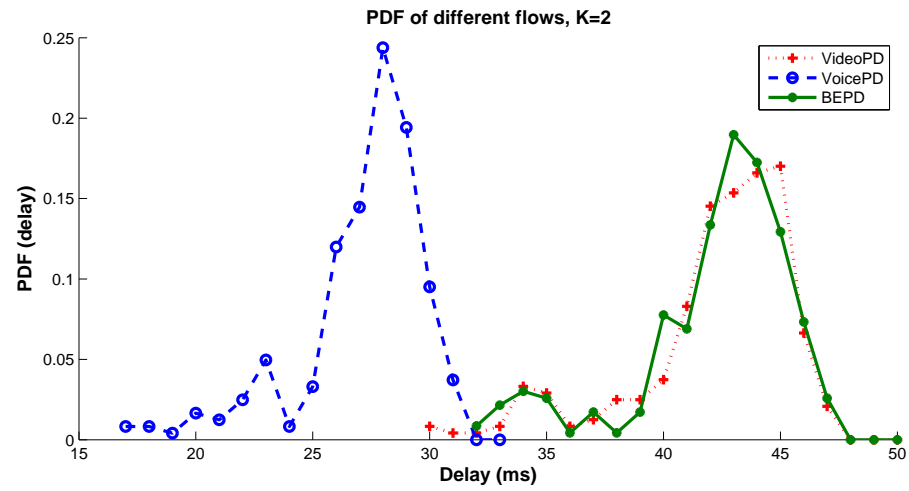


Figure 4.4: PDF of packet delay - two users in the cell

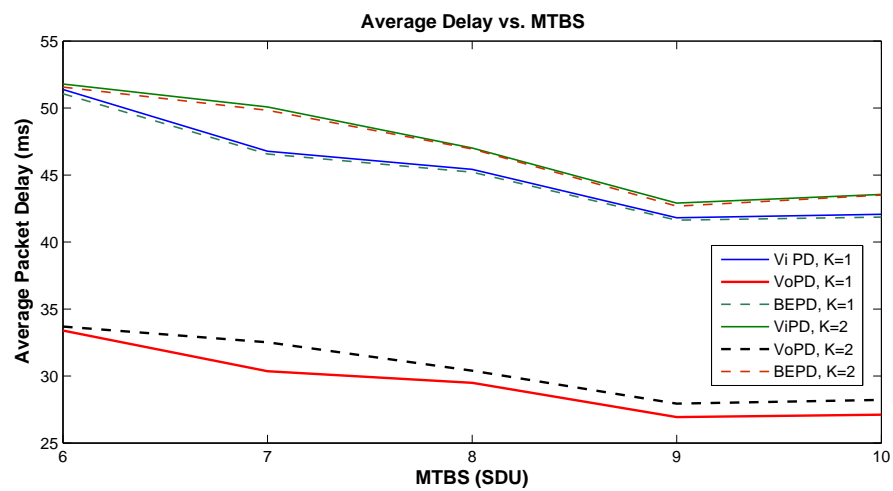


Figure 4.5: Average packet delay for different values of MTBS

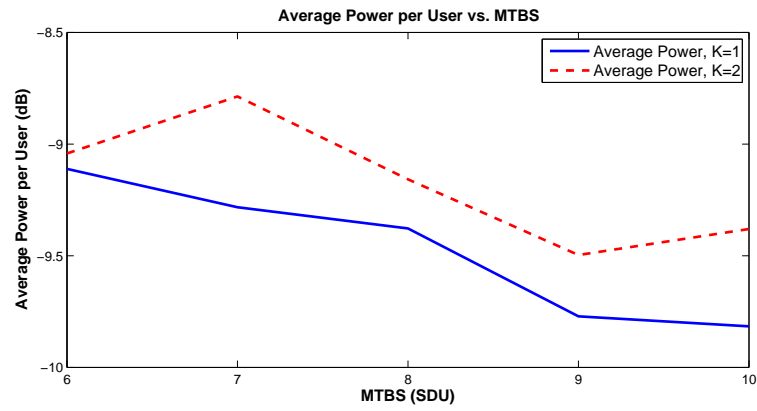


Figure 4.6: Average power delay for different values of MTBS

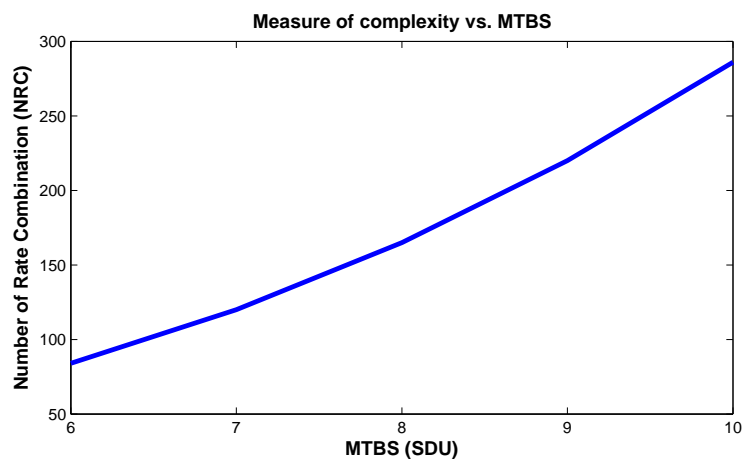


Figure 4.7: Measure of complexity vs. MTBS

general except the interval from 4 to 5 where the average power is almost stable. This situation can arise from the different simulation conditions and irregularities in the data set. Another possible reason is the No. of simulation iterations.

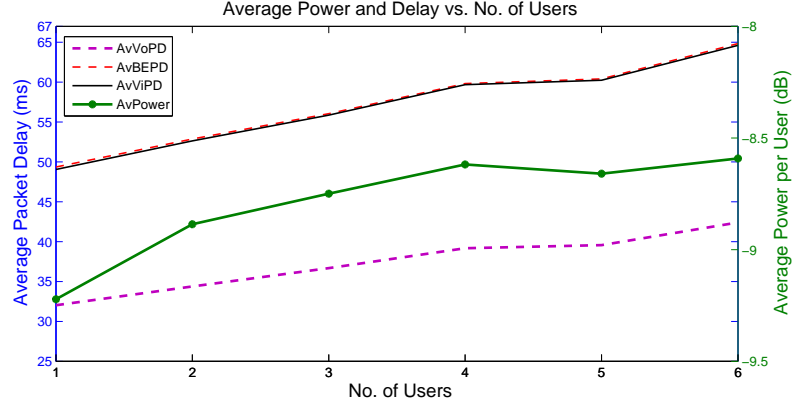


Figure 4.8: Average power and packet delay vs. No. of users

4.6 Chapter Summary

In this chapter, we have studied the uplink scheduling framework for the LTE standard. The resource allocation is divided into two tasks, and optimization techniques are used to solve the scheduling problem. Multi-objective functions are solved using a sum-weighting method and a method for specifying the weights of each objective function is suggested. Due to the difficulty of solving the problem, we relaxed some of the associated constraints. However, we solved this scheduling scheme by using a set-partitioning method. In the end, we assessed our approach in terms of average packet delays for each class of data, transmitted power per user and complexity.

Chapter 5

Adaptive Power-efficient scheduler for LTE Uplink

In this chapter, we propose an adaptive energy-saving resource allocation framework. The main goal of the proposed scheme is to prolong the battery life of mobile phones by decreasing their transmit power expenditure. The proposed scheme satisfies the users' QoS requirements and saves transmit power in an uplink LTE network. An adaptive controller is derived such that the Maximum Allowable Transmit Power (MATP) for each user changes based on the QoS factor. The proposed algorithm utilizes the Binary Integer Programming (BIP) method to solve the scheduling problem. Subsequently, a heuristic algorithm is deduced to reduce the computation time of the BIP solution. The proposed heuristic algorithm has a polynomial complexity. The proposed BIP framework and the heuristic algorithm are evaluated, and compared to the existing non-adaptive algorithm. The simulation results show that our approaches achieve power reduction and can maintain the QoS requirements.

5.1 Introduction

Smart phones have empowered users with internet access, live streaming radio, audio and video playback, navigation and much more. However, the growing functionality of smart phones results in higher data rate transmissions. High data rate requires higher transmitted energy. In uplink, the transmitters are fed from a limited battery. While battery manufacturing has not

advanced as rapidly as wireless technology, smart phone users are suffering from short battery life per charge, and thus power-efficiency transmission has become important in wireless communication. LTE is a demanding technology and provides high data rates. LTE uses SC-FDMA in uplink. SC-FDMA has lower PAPR compared with the OFDMA, which is employed in the downlink. To extend battery life, power-aware resource allocation should be employed in the uplink.

The existing works for LTE uplink scheduling can be divided into two main categories, rate adaptation and margin adaptation. Rate adaptation schedulers' objective function is to maximize the weighted-sum rate of the users. The weights are assigned by eNB to the users, for example, to impose fairness between users or to differentiate between different QoS profiles. In the SC-FDMA context, the rate adaptation objective function is subjected to three constraints:

- (a) allocation constraints: each PRB can be exclusively assigned to a single user.
- (b) contiguity constraints: to maintain low PAPR, SC-FDMA requires consecutive PRBs allocation for every user, i.e. all the assigned users' PRBs must be adjacent to each other.
- (c) maximum transmit power constraints: the user cannot transmit power higher than a threshold.

The rate adaptation problem is addressed and formulated as a binary integer problem (BIP), hence, the contiguity constraints turn the scheduling problem into non-convex optimization [39]. For instance, the MATLAB *bintprog* tool is used in [37] to solve the BIP. However, both the works [37, 39] have two main shortcomings. First, to maximize the capacity, users have to transmit at their maximum transmit power threshold, which lowers the energy efficiency of the scheduling, since the MCSs are more power-efficient at lower transmission rates [50]. Further illustration of this point will be shown in Section 5.3. Second, the schedulers do not consider the users' QoS requirements. Nevertheless, margin adaptation schedulers address the dual problem of rate adaptation, where constraints (c) are relaxed, and the objective function is to minimize the transmit power consumption. For example, Dan *et al.* followed the BIP formulation of [37] to address the margin adaptation scheduling, their objective function is

minimizing the weighted-sum power transmission subjected to constraints (a), (b), and fix rate transmission. However, to guarantee instantaneous feasible solutions, and due to the fix rate transmission, the authors of [51] did not consider the maximum transmit power in their analysis, and users were assumed to always have data to transmit.

The key contributions of this chapter are to design a novel power-efficient SC-FDMA framework and deploy a simple power threshold adjusting mechanism. The aims of this mechanism are to reduce average transmission power in the entire cell and increase the battery life time of cell phones. This approach takes into account the QoS requirements and traffic load of each user as well as channel state information, simultaneously.

5.2 System Model

In this chapter, a single cell uplink LTE transmission is considered, where K users communicate with an eNB. The whole uplink bandwidth is divided into M PRBs, and each contains 12 subcarriers. The available PRBs are assigned to the users by the eNB. Each user experiences an independent Rayleigh block-fading channel. We denote $\gamma_{i,m}$ as the instantaneous channel gain of m^{th} PRB for user i . For the Rayleigh fading channel, $\gamma_{i,m}$ follows the exponential distribution with a mean of γ_i . The effective SNR for the i_{th} user who has been assigned the continuous chunk c_i is expressed as [51]

$$\gamma_{i,eff} = \frac{P_i}{N_0 |c_i|} \sum_{m \in c_i} \frac{\gamma_{i,m}}{|c_i|} \quad (5.1)$$

where P_i denotes the transmit power, $|c_i|$ denotes the number of PRBs in the chunk c_i and N_0 is the thermal noise variance.

The granularity of the scheduler's decisions in the time domain is one LTE subframe, which is also called TTI, and one PRB in the frequency domain. Each LTE subframe contains 14 symbols, so we assumed 3 out of the 14 symbols used for signalling, during a subframe, the Transport Block (TB) is computed as

$$T_i = \lfloor 12 \times 11 \times \zeta_{mcs} \times |c_i| \rfloor \quad (5.2)$$

where ζ_{mcs} is the spectral efficiency of the used MCS, and the operator $\lfloor x \rfloor$ finds the greater integer number less than or equal to x . The LTE standard supports 15 different MCSs. In practice, link-level performance curves are used to map the SNR to BLER for each MCS. Table 5.1 shows the MCS set which is used in LTE transmission, and maps the required SNR to achieve a performance of BLER < 10% [52]. Given a chunk of PRBs assigned to a user, the effective SNR can be computed at the eNB side using Equation 5.1; accordingly, the MCS is determined based on the required BLER, and then the TB size is computed using Table 5.1.

Index	MCS	Spectral Efficiency	SNR (dB)
0	—	—	> -6.7536
1	QPSK, 78/1024	0.15237	-6.7536 : -4.9620
2	QPSK, 120/1024	0.2344	-4.9620 : -2.9601
3	QPSK, 193/1024	0.3770	-2.9601 : -1.0135
4	QPSK, 308/1024	0.6016	-1.0135 : +0.9638
5	QPSK, 449/1024	0.8770	+0.9638 : +2.8801
6	QPSK, 602/1024	1.1758	+2.8801 : +4.9185
7	16QAM, 378/1024	1.4766	+4.9185 : +6.7005
8	16QAM, 490/1024	1.9141	+6.7005 : +8.7198
9	16QAM, 616/1024	2.4063	+8.7198 : +10.515
10	64QAM, 466/1024	2.7305	+10.515 : +12.450
11	64QAM, 567/1024	3.3223	+12.450 : +14.348
12	64QAM, 666/1024	3.9023	+14.348 : +16.074
13	64QAM, 772/1024	4.5234	+16.074 : +17.877
14	64QAM, 873/1024	5.1152	+17.877 : +19.968
15	64QAM, 948/1024	5.5547	> +19.968

Table 5.1: List of MCS Indices

5.3 Adaptive Power-Efficient scheduling

The objectives of this work are to maximize the weighted-sum rates and minimize the transmit power. These two objectives contradict each other as illustrated in the following example. Consider Table 5.1, the least value of the SNR required to use MCSs numbers 5 and 10 are 0.9638 dB (1.2485), and 10.515 dB (11.2590), respectively. In terms of rate, transmitting one symbol

using MCS number 10 is equivalent to transmitting $2.7305/0.8770 = 3.1$ symbols using MSC number 5. However, transmitting 3.1 symbols using MSC number 5 consumes almost half the power (55.44%) compared to transmitting one symbol using MSC number 10, but increases the delay 3.1 times. As a conclusion, using lower transmission MCSs is energy-efficient. In contrast, higher transmission MCSs are spectral-efficient but consume higher power. To conserve power, high rate transmissions should be prevented; however, QoS requirements should be maintained.

The basic idea of the proposed scheduler is to adapt the users' Maximum Allowable Transmit Power (MATP) p_i based on users' QoS satisfaction. For each TTI, the scheduler allows users who are demanding high QoS to increase their MATP, and accordingly transmit on higher rates to meet their QoS requirements. In contrast, and to save power, MATPs are decreased for users who have low traffic loads. By adapting the MATP level, significant energy savings can be achieved.

5.3.1 Delay analysis

We assume that the users' data randomly arrive to the users' buffers with an average arrival rate of λ_i , and each user is assumed to have a buffer to store the unserved data. At n^{th} TTI, the i^{th} user's buffer status updates as follows

$$q_i[n] = q_i[n-1] + a_i[n-1] - T_i[n-1] \quad (5.3)$$

where $q_i[x]$ and $a_i[x]$ denote the queue size and arrival rate corresponding to user i and at TTI x , respectively. The LTE standard requires that probability of delay outage should be less than 2% [9].

$$P_r(d_i > D_i) \leq 2\% \quad (5.4)$$

where d_i and D_i are the head of the line delay and the delay threshold for user i , respectively. From [53], delay outage probability can be approximated as:

$$P_r(d_i > D_i) \approx \exp\left(\frac{-D_i}{E(d_i)}\right) \quad (5.5)$$

where $E(d_i)$ denotes the expected value of d_i . By substituting 5.5 in 5.4 and using Little's law, the following inequality can be derived

$$\begin{aligned} E(d_i) &= \frac{E(q_i)}{\lambda_i} \leq \frac{-D_i}{\ln(0.02)} \\ E(q_i) &\leq \Gamma_i \end{aligned} \quad (5.6)$$

where $\Gamma_i = \frac{-D_i \times \lambda_i}{\ln(0.02)}$ denotes the maximum allowable average buffer size for user i . Equation 5.6 states that: controlling the average queue size lower than Γ_i is equivalent to controlling the outage delay shown in Equation 5.4.

As a result, maintaining the average queue size $E(q_i)$ lower than Γ_i satisfies the delay requirements. This formulation is of great interest for the following reason. In LTE, only the users' queue size are available at the eNB not the actual delay, thanks to the buffer state information procedure.

5.3.2 Adaptive MATP Design and The Objective Function

In this work, a simple MATP controller is developed, where the MATP level updates every LTE frame (10 subframes) as follows

$$p'_i[l] = \begin{cases} \min(p'_i[l-1] + \delta, P_i), & \text{if } E(q_i) \geq \Gamma_i \\ p'_i[l-1] - \delta, & \text{if } E(q_i) < \Gamma_i \end{cases} \quad (5.7)$$

where $p'_i[x]$, denotes the MATP for user i at frame x , P_i denotes the user i maximum transmit power threshold, which is specified by the LTE standard, and δ is a positive constant. The objective function of the scheduler at every TTI n can be expressed as

$$\begin{aligned} &\max \sum_{i=1}^K w_i T_i \\ &\text{Subject to: constraints (a), (b), } p_i \leq p'_i \end{aligned} \quad (5.8)$$

where w_i denotes the i^{th} user weight, and constraints (a) and (b) are defined in Section 5.1.

Input: $\gamma_{i,m}$, $q_i[n]$, $p'_i[n]$

- 1: $\mathcal{PRB} = \{1, 2, \dots, M\}$, $\mathcal{PRB}_i = \emptyset$, $\forall i \in \mathcal{K}$
- 2: **while** $|\mathcal{PRB}| \neq 0$ **do**
- 3: **for** $i \in \mathcal{K}$ **do**
- 4: **if** $\mathcal{PRB}_i = \emptyset$ **then**
- 5: $C_i = \arg \max_{m \in \mathcal{PRB}} \{\gamma_{i,m}\}$
- 6: **else**
- 7: $m^* \in \{\min|\mathcal{PRB}_i| - 1, \max|\mathcal{PRB}_i| + 1\} \cap \mathcal{PRB}$
- 8: $C_i = \arg \max_{m \in m^*} \{\gamma_{i,m}\}$
- 9: **end if**
- 10: $\Delta_i = \Omega_i(\mathcal{PRB}_i \cup C_i, p'_i) - \Omega_i(\mathcal{PRB}_i, p'_i)$
- 11: **end for**
- 12: $g = \arg \max_i \{\Delta_i\}$ the greediest user
- 13: $\mathcal{PRB}_g = \mathcal{PRB}_g \cup C_g$, $\mathcal{PRB} = \mathcal{PRB} \setminus C_g$
- 14: **end while**

Table 5.2: Heuristic Allocation

5.4 Heuristic algorithm

The pseudo-code in Table 5.2 describes a heuristic algorithm to solve the scheduling problem where the operator $\Omega_i(\mathcal{PRB}_i, p'_i)$ finds the maximum TB size that can be transmitted over the PRB chunk \mathcal{PRB}_i and $p_i \leq p'_i$.

The proposed heuristic algorithm is a greedy algorithm. For each iteration, a single PRB is allocated to the greediest user as follows. Lines 4-8 find the best feasible allocation to each user considering the contiguous allocation constraints. Line 10 computes Δ_i which denotes the potential increase in TB size by adding the best feasible PRB found in Lines 4-8. Lines 12-13 determine the greediest user who achieved maximum Δ_g , then assigned the associated C_g to the greediest user, and take C_g from the unallocated PRBs set \mathcal{PRB} .

5.4.1 Complexity of the Heuristic Algorithm

The operations shown in lines 2-14 are repeated M times. In each time, at most $K \times S$ operations are required to find the best user and MCS where S is the number of used MCSs in the system. Therefore, complexity is $O(M \times S \times K)$.

5.5 Numerical evaluation

To assess the performance of our proposed algorithms, system level simulations have been conducted based on the uplink LTE model. The performance of the proposed adaptive and heuristic schedulers are compared with the non-adaptive scheduler. Table 5.3 summarizes the list of simulation parameters and assumptions. The performance of three different algorithms

Parameter	Setting
Number of PRBs	10
Channel Model	Rayleigh fading
Data Arrival Model	Poisson distribution
Average Arrival Rate	300Kbps
Number of Simulation Runs	10000
Maximum Power Transmission Threshold (P_i)	23dBm (200mw)
Power Step Size (δ)	20mw

Table 5.3: Parameter settings of the uplink LTE model

is evaluated in the following. The evaluations are observed over different average channel gain γ , and in terms of average delay, average rate, average transmission power and complexity. Figure 5.1 illustrates the averaged delay. It is expected that by increasing the average SNR, the average delay decreases. In Figure 5.1, This is correct in general except the interval from 18 to 20 dB where the average delay is almost stable. This situation can arise from the different simulation conditions and irregularities in the data set. Another possible reason is the No. of simulation iterations. However, this does not affect the general solution because for all of the SNR values the QoS requirements are met.

Figure 5.2 and Figure 5.3 present the average rate and normalized average transmission power, respectively. To explore the merits of our algorithms, we should consider these two parameters at the same time. At any given average channel gain, all of the algorithms have the same normalized rate, but the power consumption of the adaptive algorithm is less than the non-adaptive one and the heuristic algorithm is between these two. As expected, by increasing the average channel gain, the average transmitted power decreases. It is worth noting that before average channel gain γ is equal to 14 dB, the adaptive algorithm acts the same as the non-adaptive scheme due to heavy load traffic and assignment of the maximum power threshold

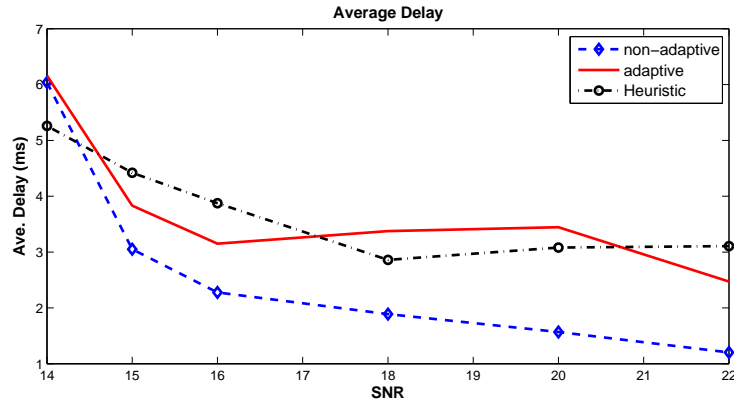


Figure 5.1: Average Delay

to satisfy users' QoS requirements.

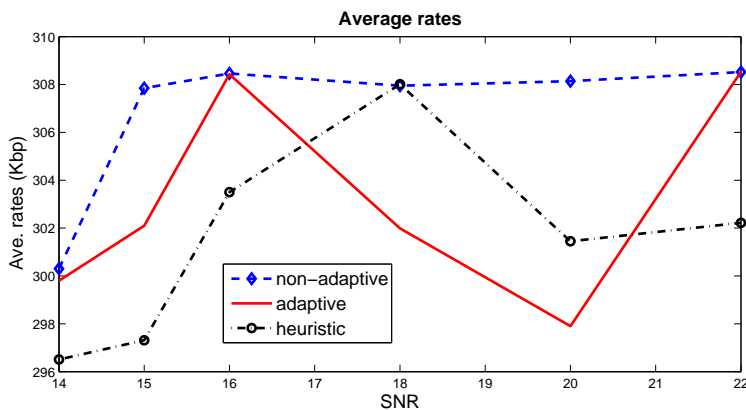


Figure 5.2: Average Rates

Figure 5.4 shows the normalized average computational time for the three schedulers relative to the base case (non-adaptive algorithm and $\gamma = 22$ dB). The MATLAB functions *tic-toc* is used as a measure of complexity over different γ . It is worth noting that our adaptive algorithm is less complex than the non-adaptive scheduler. At all times, the adaptive scheme has less or equal users' MATP (P_i) rather than the non-adaptive algorithm. Based on 5.1, the adaptive algorithm experienced less effective SNR ($\gamma_{i,eff}$) and thus lower choices in MCS selection (according to Table 5.1 per user). Reducing MCS choices causes a smaller search space and computational time compared to the non-adaptive scheduler. On the other hand, the computational time of the heuristic algorithm due to the polynomial complexity is significantly lower

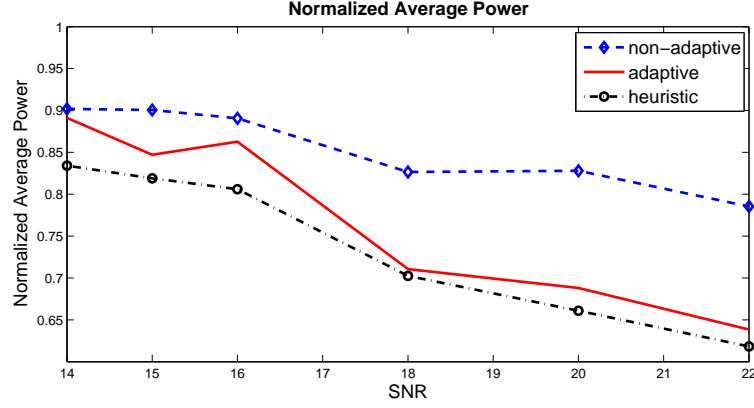


Figure 5.3: Normalized Average Power

than the adaptive algorithm, which possesses exponential complexity [39]. Consequently, our

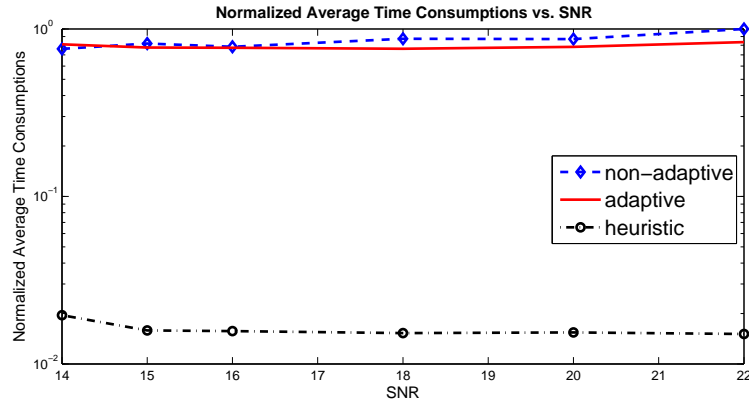


Figure 5.4: Normalized Average Time Consumptions

algorithms compared to the non-adaptive algorithm with fixed MATP not only decrease power consumption but also reduce complexity. These two goals are achieved without violating the users' QoS requirements.

5.6 Chapter Summary

In this Chapter, we propose an energy-efficient rate adaptive scheduling scheme. In order to reduce the average transmit power, a novel adaptive power threshold adjusting method is developed. In this method, the allowed power threshold is updated in each frame according to

user's QoS requirement. A heuristic method is then proposed to reduce the computational time of the algorithm. Numerical evaluations show that the adaptive algorithm and the heuristic algorithm exhibit better performance than a non-adaptive algorithm in terms of the average transmission power.

Chapter 6

Conclusion and Future Works

Energy consumption is one of the most important parameters in wireless communication. Therefore, energy efficient scheduling techniques have become more crucial in cellular network systems. In this thesis, we devised two different power-based scheduling algorithms in uplink LTE. The first algorithm focuses on power consumptions and packet delays at the same time. Three different types of data, that all of which have their own QoS requirements, are investigated in this algorithm. By using the sum-weighting method, these two different criteria are merged to an objective function. A BIP method is utilized to solve the optimization problem. In this algorithm, the effective SNR of the channel is modelled as average SNR over the assigned sub-channels for each user. Also, the least-square approximation method is employed to approximate the function that relates block error rate, desired SNR and service data rate. The performance of the algorithm is evaluated in terms of average packet delays, average power consumption per user and complexity. In the second algorithm, an adaptive energy-efficient approach is devised. Although the objective function is to maximize cell capacity, the transmitted power is limited in every LTE frame. In other words, the algorithm updates the power threshold in each scheduling decision based on the delay requirements. To decrease the complexity of the algorithm, a heuristic algorithm is presented. The numerical results show the proposed and heuristic algorithms decrease transmission power while maintaining the desired users' QoS.

6.1 Future works

There are some other scheduling factors that are not included in this study that could be interesting to investigate. These parameters are explained in Sec. 3.1. The interference estimate is not considered in this study and it would be interesting to run the simulations of this study with a model of an interference estimate and survey how the scheduling performances are affected due to cell interference. In the scheduling algorithms explained in chapters 4 and 5, HARQ retransmission can be included to make the scheduling problem close to reality. As one of the next steps, the primary attempts to implement scheduling algorithms on hardware (specially FPGA) can be studied. Another interesting issue is an exploration of the different scheduling algorithms from the processing-power-consumption aspects.

Bibliography

- [1] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, “Proportional fair frequency-domain packet scheduling for 3gpp lte uplink,” in *INFOCOM 2009, IEEE*, 2009, pp. 2611–2615.
- [2] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005. [Online]. Available: <http://books.google.ca/books?id=n-3ZZ9i0s-cC>
- [3] T. Ali-Yahiya, *Understanding LTE and its Performance*, ser. SpringerLink : Bücher. Springer, 2011. [Online]. Available: http://books.google.ca/books?id=R_oLk8yFTkkC
- [4] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian, and V. Niemi, *UMTS Networks: Architecture, Mobility and Services*. Wiley, 2005. [Online]. Available: http://books.google.ca/books?id=kX3Z9ss_5P0C
- [5] H. Holma and A. Toskala, *HSDPA/HSUPA for UMTS: high speed radio access for mobile communications*. John Wiley, 2006. [Online]. Available: <http://books.google.ca/books?id=SxcfAQAAIAAJ>
- [6] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, “Lte: the evolution of mobile broadband,” *Communications Magazine, IEEE*, 2009.
- [7] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*, ser. Wiley InterScience online books. Wiley, 2009. [Online]. Available: <http://books.google.ca/books?id=wDz07V78HZ8C>
- [8] 3rd Generation Partnership Project (3GPP) 23.401, “Technical Specification Group Services And System Aspects; GPRS enhancements for E-UTRAN access (Release 8),” 2007.
- [9] 3rd Generation Partnership Project (3GPP)23.203, “Technical Specification Group Services and System Aspects; Policy and charging control architecture.”
- [10] G. T. S. 36.213, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures.”
- [11] R. Merris, *Combinatorics*, ser. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2003. [Online]. Available: <http://books.google.ca/books?id=OM3CP4i58b4C>

- [12] J. Fan, Q. Yin, G. Li, B. Peng, and X. Zhu, "Adaptive block-level resource allocation in ofdma networks," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 11, pp. 3966–3972, 2011.
- [13] Ericsson, "System-level evaluation of OFDMfurther considerations."
- [14] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of lte networks," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, 2010, pp. 1–5.
- [15] L. Wan, S. Tsai, and M. Almgren, "A fading-insensitive performance metric for a unified link quality model," in *Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE*, vol. 4, 2006, pp. 2110–2114.
- [16] T. Jensen, S. Kant, J. Wehinger, and B. Fleury, "Fast link adaptation for mimo ofdm," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 8, pp. 3766–3778, 2010.
- [17] F. Calabrese, "Scheduling and link adaptation for uplink SC-FDMA systems," Ph.D. dissertation, Aalborg University. http://vbn.aau.dk/files/19156393/PhD_Thesis_FrancescoDavideCalabrese_final_print.pdf, 2009.
- [18] G. Monghal, K. Pedersen, I. Kovacs, and P. Mogensen, "Qos oriented time and frequency domain packet schedulers for the utran long term evolution," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, 2008, pp. 2532–2536.
- [19] D. Skoutas and A. Rouskas, "Scheduling with qos provisioning in mobile broadband wireless systems," in *Wireless Conference (EW), 2010 European*, 2010, pp. 422–428.
- [20] R. Basukala, H. Mohd Ramli, and K. Sandrasegaran, "Performance analysis of exp/pf and m-lwdf in downlink 3gpp lte system," in *Internet, 2009. AH-ICI 2009. First Asian Himalayas International Conference on*, 2009, pp. 1–5.
- [21] J. Lim, H. Myung, K. Oh, and D. Goodman, "Channel-dependent scheduling of uplink single carrier fdma systems," in *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, 2006, pp. 1–5.
- [22] T. Shi, S. Zhou, and Y. Yao, "Capacity of single carrier systems with frequency-domain equalization," in *Emerging Technologies: Frontiers of Mobile and Wireless Communication, 2004. Proceedings of the IEEE 6th Circuits and Systems Symposium on*, vol. 2, 2004, pp. 429–432 Vol.2.
- [23] G. R1-050718, "Simulation methodology for EUTRA UL: IFDMA and DFT-spread-OFDMA."
- [24] H. Myung, K. Oh, J. Lim, and D. Goodman, "Channel-dependent scheduling of an uplink sc-fdma system with imperfect channel information," in *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, 2008, pp. 1860–1864.
- [25] H. Kim, K. Kim, Y. Han, and S. Yun, "A proportional fair scheduling for multicarrier transmission systems," in *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, vol. 1, 2004, pp. 409–413 Vol. 1.

- [26] F. Calabrese, P. Michaelson, C. Rosa, M. Anas, C. Castellanos, D. Villa, K. Pedersen, and P. Mogensen, "Search-tree based uplink channel aware packet scheduling for utran lte," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, 2008, pp. 1949–1953.
- [27] F. Calabrese, C. Rosa, M. Anas, P. Michaelson, K. Pedersen, and P. Mogensen, "Adaptive transmission bandwidth based packet scheduling for lte uplink," in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, 2008, pp. 1–5.
- [28] F. Calabrese, C. Rosa, K. Pedersen, and P. Mogensen, "Performance of proportional fair frequency and time domain scheduling in lte uplink," in *Wireless Conference, 2009. EW 2009. European*, 2009, pp. 271–275.
- [29] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2619–2692, 1998.
- [30] D. M. C. R. Jain and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Systems."
- [31] E. Yaacoub, H. Al-Asadi, and Z. Dawy, "Low complexity scheduling algorithms for the lte uplink," in *Computers and Communications, 2009. ISCC 2009. IEEE Symposium on*, 2009, pp. 266–270.
- [32] C.-H. Chang, H.-L. Chao, and C.-L. Liu, "Sum throughput-improved resource allocation for lte uplink transmission," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, 2011, pp. 1–5.
- [33] D. Kim, J. Kim, H. Kim, K. Kim, and Y. Han, "An efficient scheduler for uplink single carrier fdma system," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, 2010, pp. 1348–1353.
- [34] S. Kwon and N.-H. Lee, "Uplink qos scheduling for lte system," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, 2011, pp. 1–5.
- [35] O. Delgado and B. Jaumard, "Scheduling and resource allocation in lte uplink with a delay requirement," in *Communication Networks and Services Research Conference (CNSR), 2010 Eighth Annual*, 2010, pp. 268–275.
- [36] Z. Li, C. Yin, and G. Yue, "Delay-bounded power-efficient packet scheduling for uplink systems of lte," in *Wireless Communications, Networking and Mobile Computing, 2009. WiCom '09. 5th International Conference on*, 2009, pp. 1–4.
- [37] I. Wong, O. Oteri, and W. McCoy, "Optimal resource allocation in uplink sc-fdma systems," vol. 8, no. 5, 2009, pp. 2161–2165.
- [38] F. Sokmen and T. Girici, "Uplink resource allocation algorithms for single-carrier fdma systems," in *Wireless Conference (EW), 2010 European*, 2010, pp. 339–345.

- [39] A. Ahmad and M. Assaad, "Power efficient resource allocation in uplink sc-fdma systems," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, 2011, pp. 1351–1355.
- [40] D. Dechene and A. Shami, "Energy efficient resource allocation in sc-fdma uplink with synchronous harq constraints," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–5.
- [41] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*. John Wiley & Sons, 2011.
- [42] D. Buckingham, "Information-outage analysis of finite-length codes," Ph.D. dissertation, West Virginia University, 2008.
- [43] X. Wang, G. Giannakis, and A. Marques, "A unified approach to qos-guaranteed scheduling for channel-adaptive wireless networks," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2410–2431, Dec. 2007.
- [44] Q. Liu, S. Zhou, and G. Giannakis, "Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1142–1153, May 2005.
- [45] D. J. Dechene, "Energy Efficient Reduced Complexity Multi-Service, Multi-Channel Scheduling Techniques," Ph.D. dissertation, Western University, 2011.
- [46] D. Dechene and A. Shami, "Energy efficient qos constrained scheduler for sc-fdma uplink," *In Press, Elsevier Physical Communications*.
- [47] A. Marques, L. Lopez-Ramos, G. Giannakis, J. Ramos, and A. Caamao, "Optimal cross-layer resource allocation in cellular networks using channel- and queue-state information," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 6, pp. 2789–2807, July 2012.
- [48] M. Al-Rawi, R. Jantti, J. Torsner, and M. Sagfors, "On the performance of Heuristic opportunistic scheduling in the uplink of 3G LTE networks," in *Proc. of IEEE PIMRC 2008*.
- [49] J. Cohon, *Multiobjective Programming and Planning*, ser. Dover Books on Computer Science Series. Dover Publications, 2004.
- [50] G. Miao, N. Himayat, G. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink ofdma," *Communications, IEEE Trans. on*, vol. 60, no. 1, pp. 112–120, 2012.
- [51] D. Dechene and A. Shami, "Energy-aware resource allocation strategies for lte uplink with synchronous harq constraints," *In Press, IEEE Transactions on Mobile Computing*.
- [52] J. Blumenstein, J. Ikuno, J. Prokopec, and M. Rupp, "Simulating the long term evolution uplink physical layer," in *ELMAR, 2011 Proceedings*, sept. 2011, pp. 141–144.

- [53] G. Song, Y. Li, L. Cimini, and H. Zheng, “Joint channel-aware and queue-aware data scheduling in multiple shared wireless channels,” in *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, vol. 3, 2004, pp. 1939–1944 Vol.3.

Appendix A

Source code of chapter 4

```
K=1;
M=3;
L=5; % window size to make average
beta=10000;
% average arrival data for different types of data
lambdaVideo=2000;
lambdaVoice=1000;
lambdaBestEffort= 3000;
% arrival data for different types of data
aVi= poissrnd(lambdaVideo);
aVo= poissrnd(lambdaVoice);
aBe= poissrnd(lambdaBestEffort);
%factors for weights of delays of different class of data
fVi = 0.2;
fVo =0.3;
fBe = 0.1;
% number of bits in queue
qVi=zeros(L,1);
qVo=zeros(L,1);
qBe=zeros(L,1);
for mm=1:50
    aVi= poissrnd(lambdaVideo);
    aVo= poissrnd(lambdaVoice);
    aBe= poissrnd(lambdaBestEffort);
    vtestOld=10000;
    vtest1=10000;
    x=zeros ;
    InstSNR=exprnd(10,[K, M]);
    InstSNRT=InstSNR';
    l=0;
    optOut=zeros(83,4);
```

```

for i=0:6
    for j=0:6
        for k=0:6
            if i+j+k <= 3 && i+j+k >= 1
                [x1 vtest1 a1] = Optimization(K,M,i*1000,j*1000,k*...
                1000,InstSNRT,beta,fVi,fVo,fBe,qVi,qVo,qBe,...
                lambdaVideo,lambdaVoice,lambdaBestEffort,L);
                if vtest1<vtestOld
                    vtestOld = vtest1;
                    xOld=x1;
                    TiVi=i*1000;
                    TiVo=j*1000;
                    TiBe=k*1000;
                end
                l=l+1;
                optOut(1,1)=vtest1;
                optOut(1,2)=i;
                optOut(1,3)=j;
                optOut(1,4)=k;
            end
        end
    end
end
% update the values of queue
for tt=1:9
    qVi(tt)=qVi(tt+1);
    qVo(tt)=qVo(tt+1);
    qBe(tt)=qBe(tt+1);
end
qVi(10)=qVi(9)+ aVi - TiVi;
qVo(10)= qVo(9)+ aVo - TiVo;
qBe(10)= qBe(9)+ aBe - TiBe;
end
aVi= poissrnd(lambdaVideo);
aVo= poissrnd(lambdaVoice);
aBe= poissrnd(lambdaBestEffort);
vtestOld=10000;
vtest1=10000;
x=zeros;
InstSNR=exprnd(10,[K, M]);
InstSNRT=InstSNR';
l=0;
optOut=zeros(83,4)
for i=0:6
    for j=0:6

```



```

Nmin = 1;
Nmax = M;
% calculating C = number of columns in A matrix
S= Nmax - Nmin +1;
Z=0;
for i=0 : S-1
    Z= Z+Nmin+i;
end
C=S*(M+1)- Z;
A1= zeros (M,C);
%for x Rbs (has Ml columns)
preColumn = 0;
for x=Nmin: Nmax;
    Ml= M-x+1;
    B =zeros (M,Ml);
    D =zeros (M,Ml);
    for j=1:Ml
        for i=1:M
            if (i-j>=0) && (i-j<x)
                B(i ,j)=1;
            end
        end
    end
    D(1 : M , 1 :Ml )= B;
    y = (Ml+preColumn);
    A1(:,preColumn+1: y)= D;
    preColumn = preColumn+Ml;
end
% for all users the matrix is the same
A = zeros (M,C*K) ;
preCol =0 ;
for k = 1 : K
    A(:,(k-1)*C+1: k*C )=A1;
end
%% generate Aeq matrix
Aeq = zeros (K,C*K);
smallOnes = ones(1, C);
for k=1:K
    Aeq(k,(k-1)*C+1: k*C )= smallOnes;
end
%% generate cost matrix
LSAMP1=[1.1748 5.2471 1.1019; 1.1208 2.624 1.0723; 1.0977 1.7495...
1.0591; 1.0841 1.3122 1.0512; 1.0749 1.0498 1.0458;1.0682 0.8749 1.0418]
LSAMP2=[1.063 0.7499 1.0387; 1.0588 0.6562 1.0362; 1.0553 0.5833 ...
1.0342; 1.0524 0.525 1.0324; 1.0499 0.4772 1.0309; 1.0478 0.4375 1.0296]

```

```

LSAMP3=[1.0459 0.4038 1.0284; 1.0441 0.375 1.0274; 1.0426 0.35 1.0265;...
1.0412 0.3281 1.0256; 1.04 0.3088 1.0249; 1.0388 0.2917 1.0242];
LSAMP4=[1.0378 0.2763 1.0235; 1.0368 0.2625 1.0229; 1.0359 0.25 1.0224;...
1.0351 0.2386 1.0219; 1.0343 0.2283 1.0214; 1.0336 0.2188 1.0209];
LSAMP =[LSAMP1;LSAMP2;LSAMP3;LSAMP4];
Cost=zeros(K*C,1);
CostdB=zeros(K*C,1);
measEffSNR=zeros(K*C,1);
reqEffSNR=zeros(K*C,1);
% generate reqEffSNRdB and measEffSNR for all user
for k = 1 : K % user k
    for i = 1 : C
        Ni=nnz(A(:,((k-1)*C)+i));
        reqEffSNR(((k-1)*C)+i,1) = (LSAMP(Ni,1)*exp(LSAMP(Ni,2)*1e-3...
            *TiFixed)-LSAMP(Ni,3));
        measEffSNR(((k-1)*C)+i,1) =sum( (InstSNRT(:,k)).*(A(:,((k-1)...
            *C)+i)))/(Ni*Ni);
    end
end
QIVi= (sum(qVi)+TiVi)/(L*lambdaVideo);
QIVo= (sum(qVo)+TiVo)/(L*lambdaVoice);
QIBe= (sum(qBe)+TiBe)/(L*lambdaBestEffort);
Cost = reqEffSNR ./ measEffSNR; %power
delayPart= beta*(fVi*QIVi+fVo*QIVo+fBe*QIBe);
totCost=Cost+delayPart;
CostdB = 10*log10(Cost);
%% generate b and beq matrix
b=ones(M,1);
beq=ones(K,1);
%% solve the problem
[x,fval,exitflag] = bintprog(totCost,A,b,Aeq,beq);

```

Appendix B

Source code of chapter 5

```
function [Delay_all , Rate_all , Power_all , Pct]= M_rate_ctrl(K, It ,M, Pth ,...
    SNR1, lambda1 ,GBR1,Dm1,RBm1,L1 ,SNR2, lambda2 ,GBR2,Dm2,RBm2,L2 ,...
    SNR3, lambda3 ,GBR3,Dm3,RBm3,L3 ,SNR4, lambda4 ,GBR4,Dm4,RBm4,L4 ,...
    extra_TTI , D_marg ,ch ,NGBR, final , step_size )
D_des1=-Dm1/(log(.02));% desire value
Buffers1=D_des1;
Pct=[];
D_des2=-Dm2/(log(.02));% desire value
Buffers2=D_des2;
D_des3=-Dm3/(log(.02));% desire value
Buffers3=D_des3;
D_des4=-Dm4/(log(.02));
Buffers4=D_des4;
rate1= repmat(GBR1,L1,1); Delay1= repmat(D_des1,L1,1); rate2= repmat(GBR2,...
    L2,1); Delay2= repmat(D_des2,L2,1);
rate3= repmat(GBR3,L3,1); Delay3= repmat(D_des3,L3,1); rate4= repmat(GBR4,...
    ,L4,1); Delay4= repmat(D_des4,L4,1);
no_data=0;
P_TH1=Pth;
P_TH2=Pth;
P_TH3=Pth;
P_TH4=Pth;
pt=[];
for t=1:It+extra_TTI
    if t>=It
        no_data=1;
    end
    if t<It-final
        if mod(t,L1/4)==0;
            P_TH1 = P_ctrl(P_TH1, mean(Delay1(t+L1-1,1)),D_des1(1),Pth,...
                step_size);
```

```

        P_TH2 = P_ctrl(P_TH2, mean(Delay2(t+L1-1,1)), D_des2(1), Pth, ...
            step_size);
    end
    else
        P_TH1 = Pth;
        P_TH2 = Pth;
        P_TH3 = Pth;
        P_TH4 = Pth;
    end
    pct(t)=P_TH1;
    ch_t1=ch(:,t,1);
    [T_i_t1, P_c1, A1, Buffers1x]=main_Mr(lambda1, GBR1, RBm1, ch_t1, M, P_TH1, ...
        Buffers1, no_data);
    ch_t2=ch(:,t,2);
    [T_i_t2, P_c2, A2, Buffers2x]=main_Mr(lambda2, GBR2, RBm2, ch_t2, M, P_TH2, ...
        Buffers2, no_data);
    ch_t3=exprnd(SNR3,M,1);
    [T_i_t3, P_c3, A3, Buffers3]=main_Mr(lambda3, GBR3, RBm3, ch_t3, M, P_TH3, ...
        Buffers3, no_data);
    ch_t4=exprnd(SNR4,M,1);
    [T_i_t4, P_c4, A4, Buffers4]=main_Mr(lambda4, GBR4, RBm4, ch_t4, M, P_TH4, ...
        Buffers4, no_data);
    if t<It
        Buffers1=Buffers1+GBR1+NGBR(:,t,1)';
        Buffers2=Buffers2+GBR2+NGBR(:,t,2)';
    end
    switch K
        case 1,
            A=A1; s_A1=size(A1);
            Aeq1=[ones(1,s_A1(2))]; Aeq=[Aeq1];
            T_i_t=[T_i_t1];
            b=ones(M,1); beq=ones(K,1);
            options =optimset('LargeScale','on','Simplex','on','Display'...
                , 'off','TolXInteger',1e-16);
            [x fval]= bintprog(-T_i_t,A,b,Aeq,beq,[], options);
            Sol=find(x);
            P_c=[P_c1'];
            transmit1=T_i_t(Sol(1),:);    transmit2=0;    transmit3=0;
            transmit4=0;
        case 2,
            A=[A1 A2]; s_A1=size(A1);    s_A2=size(A2);
            Aeq1=[ones(1,s_A1(2)) zeros(1,s_A2(2))];
            Aeq2=[zeros(1,s_A1(2)) ones(1,s_A2(2))];
            Aeq=[Aeq1; Aeq2];
            T_i_t=[T_i_t1; T_i_t2];

```

```

b=ones(M,1);      beq=ones(K,1);
options =optimset('LargeScale','on','Simplex','on',...
    'Display','off','TolXInteger',1e-16);
[x fval]= bintprog(-T_i_t,A,b,Aeq,beq,[],options);
Sol=find(x);
P_c=[P_c1' P_c2'];
transmit1=T_i_t(Sol(1),:);    transmit2=T_i_t(Sol(2),:);
transmit3=0;      transmit4=0;
case 3,
A=[A1 A2 A3];
s_A1=size(A1);    s_A2=size(A2);    s_A3=size(A3);
Aeq1=[ones(1,s_A1(2)) zeros(1,s_A2(2)) zeros(1,s_A3(2))];
Aeq2=[zeros(1,s_A1(2)) ones(1,s_A2(2)) zeros(1,s_A3(2))];
Aeq3=[zeros(1,s_A1(2)) zeros(1,s_A2(2)) ones(1,s_A3(2))];
Aeq=[Aeq1;Aeq2;Aeq3];
T_i_t=[T_i_t1;T_i_t2;T_i_t3];
b=ones(M,1);      beq=ones(K,1);
options =optimset('LargeScale','on','Simplex','on',...
    'Display','off','TolXInteger',1e-16);
[x fval]= bintprog(-T_i_t,A,b,Aeq,beq,[],options);
Sol=find(x);
P_c=[P_c1' P_c2' P_c3'];
transmit1=T_i_t(Sol(1),:);    transmit2=T_i_t(Sol(2),:);...
    transmit3=T_i_t(Sol(3),:);    transmit4=0;
case 4,
A=[A1 A2 A3 A4]; s_A1=size(A1);    s_A2=size(A2);...
    s_A3=size(A3);    s_A4=size(A4);% Sizes
Aeq1=[ones(1,s_A1(2)) zeros(1,s_A2(2)) zeros(1,s_A3(2))...
    zeros(1,s_A4(2))];
Aeq2=[zeros(1,s_A1(2)) ones(1,s_A2(2)) zeros(1,s_A3(2))...
    zeros(1,s_A4(2))];
Aeq3=[zeros(1,s_A1(2)) zeros(1,s_A2(2)) ones(1,s_A3(2))...
    zeros(1,s_A4(2))];
Aeq4=[zeros(1,s_A1(2)) zeros(1,s_A2(2)) zeros(1,s_A3(2))...
    ones(1,s_A4(2))];
Aeq=[Aeq1;Aeq2;Aeq3;Aeq4];
T_i_t=[T_i_t1;T_i_t2;T_i_t3;T_i_t4];
b=ones(M,1);      beq=ones(K,1);
options =optimset('LargeScale','on','Simplex','on',...
    'Display','off','TolXInteger',1e-16);
[x fval]= bintprog(-T_i_t,A,b,Aeq,beq,[],options);
Sol=find(x);
P_c=[P_c1' P_c2' P_c3' P_c4'];
transmit1=T_i_t(Sol(1),:);    transmit2=T_i_t(Sol(2),:);
transmit3=T_i_t(Sol(3),:);    transmit4=T_i_t(Sol(4),:);

```

```

end
if transmit1 ~= 0;
    shar1=floor( transmit1 *(1e-8+Buffers1 )/sum(1e-8+Buffers1 ));
    P1(t)=P_c( Sol(1));
else
    P1(t)=0;
    shar1=[0 0 0 0];
end
if transmit2 ~= 0;
    shar2=floor( transmit2 *(1e-8+Buffers2 )/sum(1e-8+Buffers2 ));
    P2(t)=P_c( Sol(2));
else
    P2(t)=0;
    shar2=[0 0 0 0];
end
if transmit3 ~= 0;
    shar3=floor( transmit3 *(1e-8+Buffers3 )/sum(1e-8+Buffers3 ));
    P3(t)=P_c( Sol(3));
else
    P3(t)=0;
    shar3=[0 0 0 0];
end
if transmit4 ~= 0;
    shar4=floor( transmit4 *(1e-8+Buffers4 )/sum(1e-8+Buffers4 ));
    P4(t)=P_c( Sol(4));
else
    P4(t)=0;
    shar4=[0 0 0 0];
end
%% OUTPUT
Buffers1=Buffers1 - shar1;    Buffers2=Buffers2 - shar2;
Buffers3=Buffers3 - shar3;    Buffers4=Buffers4 - shar4;
Delay1=[Delay1; Buffers1];    Delay2=[Delay2; Buffers2];
Delay3=[Delay3; Buffers3];    Delay4=[Delay4; Buffers4];
rate1=[rate1; shar1];         rate2=[rate2; shar2];
rate3=[rate3; shar3];         rate4=[rate4; shar4];
pt=[pt P_TH1];
end
Delay1=Delay1(L1+1:L1+It ,:);    Delay2=Delay2(L2+1:L2+It ,:);
Delay3=Delay3(L3+1:L3+It ,:);    Delay4=Delay4(L4+1:L4+It ,:);
rate1=rate1(L1+1:L1+It ,:);      rate2=rate2(L2+1:L2+It ,:);
rate3=rate3(L3+1:L3+It ,:);      rate4=rate4(L4+1:L4+It ,:);
Delay_all=zeros( It ,4,K);
Rate_all=zeros( It ,4,K);
Delay_all1=[Delay1 ,Delay2 ,Delay3 ,Delay4 ];

```

```

Rate_all1=[rate1 ,rate2 ,rate3 ,rate4 ];
% Normalized %
D_des_all=[(1+D_marg)*D_des1;(1+D_marg)*D_des2;(1+D_marg)*D_des3 ;...
(1+D_marg)*D_des4 ];
Rate_norm=[GBR1+lambda1+1e-8;GBR2+lambda2+1e-8;GBR3+lambda3+1e-8 ;...
GBR4+lambda4+1e-8];
Power_all1=[P1' ,P2' ,P3' ,P4' ]/ Pth ;
for k=1:K
    for n=1:4
        Delay_all (: ,n,k)=Delay_all1 (: ,n+4*(k-1));
        %Rate_all (: ,n,k)=Rate_all1 (: ,n+4*(k-1))/Rate_norm(k,n);
        Rate_all (: ,n,k)=Rate_all1 (: ,n+4*(k-1));
        Power_all (: ,k)=Power_all1 (: ,k);
    end
Delay_all=Delay_all (: ,1 ,k);
Rate_all=Rate_all (: ,1 ,k);
end

```

Curriculum Vitae

Name: Aidin ReyhaniMasoleh

Post-Secondary Education and Degrees: 2001 - 2006 B.Sc.
Electrical Engineering
Iran University of Science and Technology
Tehran, Iran

2011 - 2013 M.E.Sc.
Electrical and Computer Engineering
University of Western Ontario
London, ON

Related Work Experience: Teaching Assistant
The University of Western Ontario
2012 - 2013

Publications:

- [1] "Heterogeneous Delay-Power Resource Allocation in Uplink LTE" Wireless and Mobile Networking Conference (WMNC) , Dubai, UAE, April 2013.
- [2] "Adaptive Power-Efficient Scheduler for LTE Uplink" prepared for submission.