UPLINK SCHEDULING AND RESOURCE ALLOCATION SCHEMES FOR LTE-ADVANCED SYSTEMS THAT INCORPORATE RELAYS OR CARRY HETEROGENEOUS TRAFFIC

by

Rukhsana Ruby

B.Sc. Bangladesh University of Engineering & Technology, 2004
M.Sc. University of Victoria, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2015

Abstract

Scheduling and resource allocation is one of the important tasks of the radio resource management layer in long term evolution (LTE) and LTE-Advanced wireless systems. Uplink scheduling and resource allocation is considered more challenging compared to the downlink case because of individual users' power constraints and the discrete nature of spectrum assignment. Downlink scheduling and resource allocation has extensively been studied for relay equipped or heterogeneous traffic networks, but less work has been considered for the LTE uplink case. In this thesis, we have proposed a few uplink scheduling and resource allocation schemes of LTE-Advanced systems that incorporate relay(s) or carry heterogeneous traffic.

First, we have proposed a basic uplink scheduling and resource allocation scheme for decode-and-forward relay aided systems. Existing uplink scheduling works have looked at the problem from different angles instead of basic scheduling and resource allocation. In addition of having optimal resource allocation, the proposed scheme is adaptive. If the system has some bad or redundant relays, the proposed scheme can detect and recommends them to be deactivated.

Having observed the difficulty in deciding which users to serve for a relay under the constraint of limited power, second, we have proposed a joint source and relay power allocation scheme for an amplify-and-forward relayed system. Existing works of this problem have ignored one term in their problem formulation, and hence failed to offer the optimal solution for all possible scenarios. We have taken care of that missing term in our work, and have shown the performance improvement comparing with the existing works. In this solution, all entities in the network work in an altruistic manner towards maximizing the network capacity. However, in the real world, the nodes may want some benefits while sacrificing their resource. To model the selfish behavior of the nodes, in the third work, we have proposed a game theoretical solution of this problem.

Fourth, we have proposed an uplink scheduling and resource allocation scheme for a network which carries heterogeneous traffic. Although there are some existing uplink scheduling works dealing QoS in heterogeneous traffic networks, those were not careful about detailed standard specific all constraints. In addition to meet the conflicting requirements of QoS for different traffic, the proposed scheme takes the resource utilization constraint into account which is designed to benefit the network operators.

Preface

Many of the results that are presented in this thesis have been published in academic journals or presented at academic conferences:

- Rukhsana Ruby and Victor C.M. Leung, "Uplink scheduling solution for enhancing throughput and fairness in relayed long-term evolution networks," *IET Communications*, Volume 8, Issue 6, 17 April 2014, p. 813–825. (Linked to Chapter 3)

- Rukhsana Ruby, Victor C.M. Leung and David G. Michelson, "Uplink Scheduler for SC-FDMA based Heterogeneous Traffic Networks with QoS Assurance and Guaranteed Resource Utilization," *IEEE Transaction on Vehicular Technology*, 2014, p. 1–17. (Linked to Chapter 6)

- Rukhsana Ruby, Victor C.M. Leung and David G. Michelson, "Centralized and Game Theoretical Solutions of Joint Source and Relay Power Allocation for AF Relay based Network," *IEEE Transactions on Communications*, 2015, p. 1–16. (Linked to Chapters 4 and 5)

- Rukhsana Ruby, Amr Mohamed and Victor C.M. Leung, "Utility-Based Uplink Scheduling Algorithm for Enhancing Throughput and Fairness in Relayed LTE Networks," *IEEE LCN*, 2010: 520-527. (Linked to Chapter 3)

- Rukhsana Ruby and Victor C.M. Leung, "Optimal Edge Nodes Selection and

Power Allocation in Relay Enabled Network," *IEEE CCECE*, 2011: 1347-1350. (Linked to Chapter 4)

- Rukhsana Ruby and Victor C.M. Leung, "Towards QoS assurance with revenue maximization of LTE Uplink Scheduling," *IEEE CNSR*, 2010: 202-209. (Linked to Chapter 6)

I defined the research problems considered in this thesis and am the principal contributor to all works presented here. Prof. Leung and Prof. Michelson provided editorial feedback concerning the entire thesis and provided specific guidance and feedback concerning Chapters 1, 2 and 7. Chapters 3, 4, 5 and 6 are based on the journal manuscripts listed above. Although the work on Chapter 3 began with the conference paper that I wrote with Dr. Amr Mohamed, the problem formulation and solution approach in the journal paper is significantly different.

Table of Contents

Appendices

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

1G	First Generation Cellular
2G	Second Generation Cellular
3G	Third Generation Cellular
3GPP	Third Generation Partnership Project
4G	Fourth Generation Cellular
AF	Amplify and Forward
AMR	Adaptive Multirate Codec
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CDMA	Code Division Multiple Access
CF	Compress and Forward
CSI	Channel State Information
DF	Decode and Forward
E-UTRA	Evolved UMTS Terrestrial Radio Access
EXP	Exponential
FD	Full Duplex
FTP	File Transfer Protocol
GP	Geometric Programming
HD	Half Duplex
HoL	Head of Line

HSDPA	High Speed Downlink Packet Access
IFDMA	Interleaved Frequency Division Multiple Access
i.i.d	Independent and Identically Distributed
IP	Internet Protocol
KKT	Karush-Kuhn-Tucker
LFDMA	Localized Frequency Division Multiple Access
LTE	Long Term Evolution
LWDF	Largest Weighted Delay First
MAC	Medium Access Control
M-LWDF	Modified LWDF
MT	Maximum Throughput
MW	Max Weight
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PAPR	Peak to Average Power Ratio
PDU	Protocol Data Unit
PF	Proportionally Fair
QCI	Quality of Service Class Indicator
QoS	Quality of Service
RB	Resource Block
RLC	Radio Link Control
RTP	Real Time Protocol
SC-FDMA	Single Carrier Frequency Division Multiple Access
SE	Stackelberg Equilibrium
SINR	Signal to Interference and Noise Ratio

SNR	Signal to Noise Ratio
TCP	Transmission Control Protocol
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications System
VoIP	Voice over IP

Mathematical Notations

$E(\cdot)$	Statistical expectation.
$ln(x)$	Logarithmic of x with base 2.
e^x	Exponential function of x.
H_{dB}	H is a number in decibel unit.
$\sum$	Sum operator.
$\prod$	Multiplication operator.
$\sqrt{x}$	Square root of x.
x^{-1}	Inverse of x, 1/x.
$[x]^+$	Non–negative value of x, $[x]^+ = \max\{x, 0\}$.
$\|x\|$	Absolute value of x.
$x{\pm}y$	Either $x + y$ or $x - y$.
$x{\mp}y$	Either $x - y$ or $x + y$.
$\lfloor x \rfloor$	Floor operation on x, largest integer value not greater than x.
$\lceil x \rceil$	Ceiling operation on x, largest integer value smaller than x.
$a?b : c$	if expression a is true then return b, else return c.
$O(\cdot)$	Order of a function, $f(x)$ is $O(x)$ if $\lim_{x \to 0} f(x)/x = 0$.
$\mathbf{M}^T$	Transpose of matrix $\mathbf{M}$.
$\frac{\partial H}{\partial x}$	Partial derivative of function H with respect to x.
$\frac{\partial^2 H}{\partial x^2}$	Partial derivative of function $\frac{\partial H}{\partial x}$ with respect to x.
$\frac{\partial^2 H}{\partial x \partial y}$	Partial derivative of function $\frac{\partial H}{\partial x}$ with respect to y.

$x \in \mathbf{X}$	x is a variable which points to an element of set $\mathbf{X}$.
$\forall x \in \mathbf{X}$	x can be any element of set $\mathbf{X}$.
$\exists x \in \mathbf{X}$	x is some element of set $\mathbf{X}$, not necessarily points to any element of $\mathbf{X}$.
$[\mathbf{X}\|y]$	This is a set which contains all elements of set $\mathbf{X}$ except y.
$[x \in \mathbf{X}\|x{\neq}y]$	x points to an element of $\mathbf{X}$ given that $x{\neq}y$.
$\mathbb{R}$	The set of real numbers.
$[x, y)$	The set which contains all real numbers in between x and y, including x and excluding y.
$\mathbf{A}{\cup}\mathbf{B}$	Union of sets $\mathbf{A}$ and $\mathbf{B}$.
$\mathbb{R}^n$	A vector which has n elements, and each element is from set $\mathbb{R}$.

Acknowledgments

First of all, I am grateful to Almighty GOD for giving me courage, strength and making the way towards pursuing this degree. HE gave me energy time to time to get up when I was about to fall down during the critical periods of my life.

Second, I am thankful to my supervisor Prof. Victor C.M. Leung for supporting me financially some parts of the program. I appreciate the help of my co-supervisor Prof. David G. Michelson for his assistance in the final year.

Insightful comments by my committee members Prof. Vikram Krishnamurthy and Prof. Jane Wang helped me to materialize all problems in the thesis. They supported the ideas of my PhD proposal at the beginning, and finally the thesis. I am thankful to them as well.

Special thanks to WiNMoS group members Dr. Jun Bae Seo, Dr. Haoming Li, Dr. Hu Jin, Javad Hajipour, Hasen Nisanfar, Xin Ge and others for giving me feedback when I presented my works in our private group meetings. I have some collaborators in my group with whom I have some works as a co-author. They are Li Zhou, Xuan Dong, Saad Mahboob, Shiguo Wang, Javad Hajipour and Arghavan Emami. They made my time worthy and enjoyable, and I would like to acknowledge them. Many thanks to my friend Dr. Yanyan Zhuang and my housemate Victoria Karner to get along with me during my loitering time.

Finally, I am immensely indebted to my family, i.e., my mother and sisters for their love and blessings. They were my inspiration to live in this planet. This thesis is dedicated to the loving memory of my father.

Chapter 1

Introduction

Cellular networks as well as wireless services have undergone unprecedented development over the last few decades. With the increasing demand of subscribers, handheld devices now support sophisticated applications, such as interactive video, online games, etc. In order to make these applications work with the subscribers at the minimum quality of service (QoS), network operators need very high data rate transmission media. Accordingly, digital signal processing techniques, microprocessor technology, and radio frequency (RF) engineering have made a considerable advancement, and thus cellular network technologies are emerging, such as Long Term Evolution (LTE) and LTE-Advanced. In general, the performance of cellular networks is based predominantly on its achievable throughput, i.e., how many data bits can be conveyed successfully in a given amount of time; average end to end delay of the data packets, i.e., on average how much time is required to transmit a data packet; fairness, i.e., do the subscribers experience the same QoS regardless of their distance from the base station, etc. Figure 1.1 shows different types of emerging applications and their QoS requirements in terms of throughput and latency.

Conventionally, the performance of such systems can be improved by increasing the channel bandwidth and/or transmit power. However, the amount of channel bandwidth is limited due to the physical limitation, and thus cannot be exploited lavishly. On the other hand, there is great need to conserve power when the transmitter is battery operated as in the case of cellular handsets. A feasible response to these

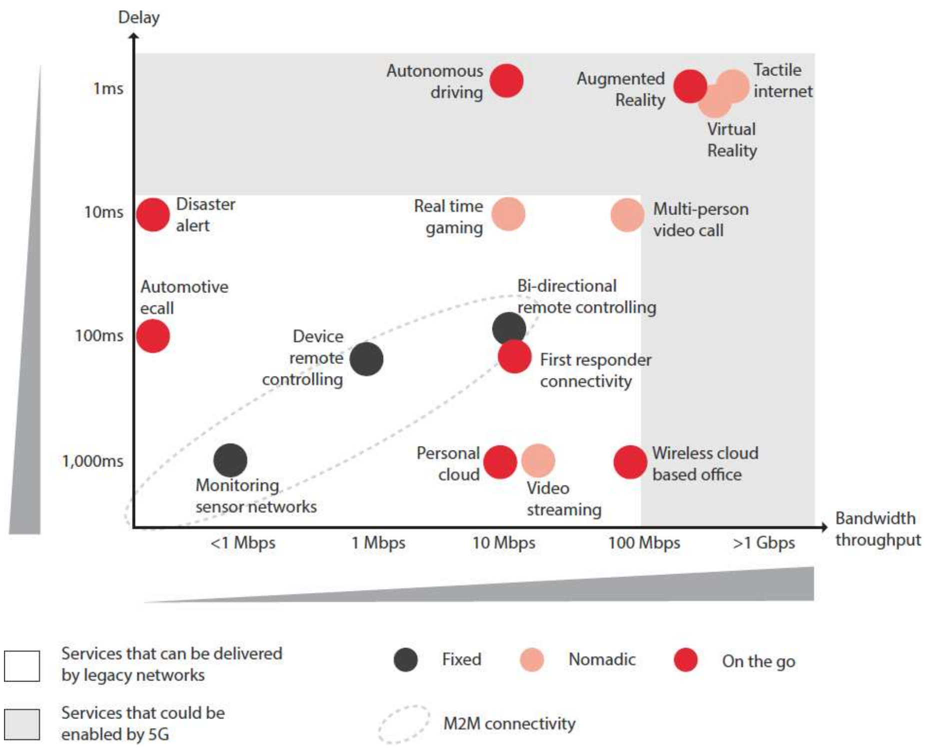


Figure 1.1: Bandwidth and latency requirements of potential applications [1].

issues is to design the system in such a way that these resources are used efficiently and effectively. While making the design of such systems, several optimization problems need to be solved in order to produce suitable scheduling and resource allocation schemes. However, depending on the transmission direction, the formulation of the scheduling and resource allocation problems, and the corresponding solutions may vary. For example, the formulation of the downlink scheduling problem is different compared to the uplink one because of the varying nature of constraints. Moreover, for the uplink scheduling, physical channels may need to be assigned among the subscribers in a certain manner compared to the downlink one. These problems make further difference if the system is equipped with digital relays or the system is loaded with the subscribers running different applications with distinct QoS.

In this thesis, we have proposed a few uplink scheduling and resource allocation schemes that improve the performance of LTE-Advanced systems that incorporate relays or carry heterogeneous traffic. In order to recognize the contributions of this thesis, the readers are encouraged to go through Chapter 2. In addition to related works in the context of this thesis, we have briefly summarized the evolution of LTE-Advanced systems and their radio resource management policies in Chapter 2. Moreover, since the scheduling concept in this thesis is based on the uplink spectrum-users mapping, we have briefly reviewed the uplink spectrum access mechanism there.

The remainder of this chapter is organized as follows. We briefly summarize the contributions of this thesis in Section 1.1, while Section 1.2 describes its structure.

1.1 Research Problems and Contributions

In the LTE uplink, the time-frequency resource unit is the resource block (RB), defined as 7 (occasionally 6) consecutive orthogonal frequency division multiplexing

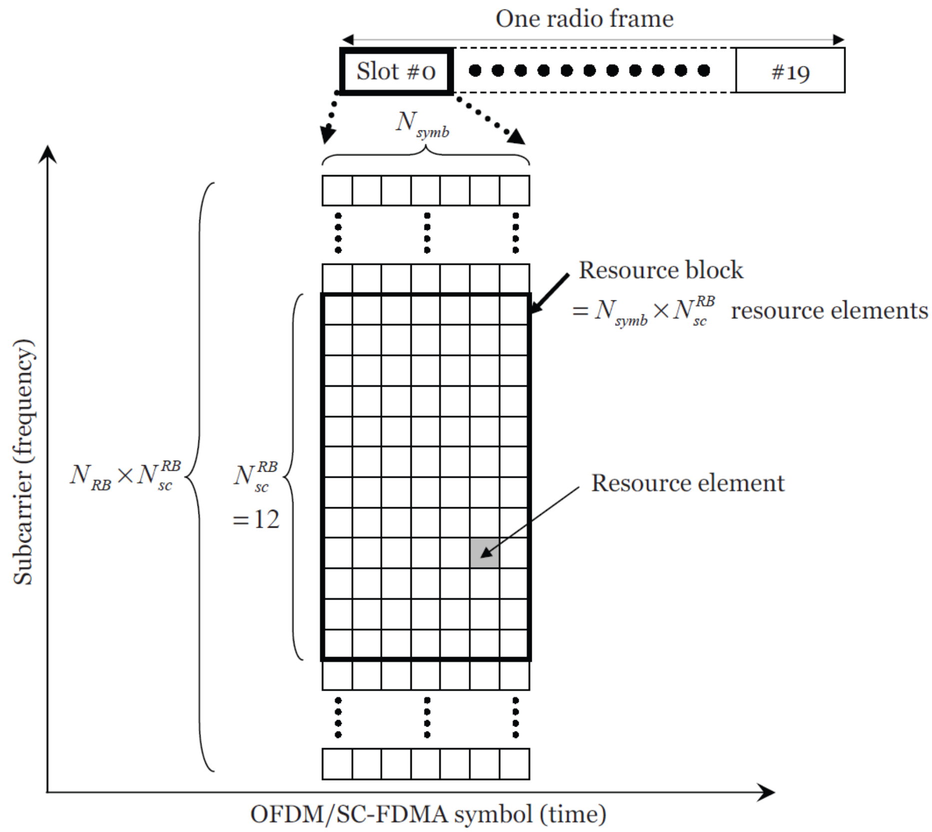


Figure 1.2: LTE uplink resource grid for one slot [2].

(OFDM) or single carrier frequency division multiple access (SC-FDMA) symbols in the time domain and 12 consecutive subcarriers in the frequency domain. The physical description of RBs is shown in Figure 1.2. The resources we have considered in this thesis are RBs and power. At the minimal transmission time interval (TTI) of the LTE frame structure, the available number of RBs is fixed and constant. The detailed description of LTE frame and channel structure is given in Chapter 2. The basic scheduling and resource allocation is to select a set of users, determining RB-user mapping and setting of transmit power on each RB in each TTI.

- **Contribution 1:** First, we have proposed a basic uplink scheduling and resource allocation scheme for a LTE network which has some pre-configured static decode-and-forward (DF) relays with full duplex (FD) functionality. As mentioned in Chapter 2 that there were extensive works on the basic downlink scheduling for relay equipped LTE networks taking various factors, such as fairness, quality of service (QoS), spatial reuse into account. There were some contemporary uplink scheduling works as well, however those looked at the scheduling problem from different angles instead of giving solution for the basic uplink scheduling and resource allocation. The basic uplink scheduling and resource allocation is necessary to measure the system performance precisely in the granular level. To solve the problem, first, we have formulated the problem assuming the routing information of all nodes are pre-configured. Dual decomposition method is usually used to solve this type of problem. Having noticed the high computational complexity of this solution, we have proposed a set of suboptimal scheduling schemes. Deploying a large number of relays may not be useful to the basic user nodes, and hence the proposed schemes are adaptive and have ability to distinguish useful relays from the not-useful ones. Besides, if the relays are in the bad locations, sometimes, the proposed schemes can distinguish those bad relays. Through extensive simulation, we have shown the performance improvement of our scheduling schemes comparing with other existing work in the literature.

- **Contribution 2:** In the first work, we have seen, if a relay needs to help many sources, even if the assignment of RBs are known, optimal power allocation among the relay and sources is a difficult problem. There are many existing works in this context no matter the relay has DF or amplify-and-forward (AF) functionality. Although the performance of a DF relay is better, researchers often recommend AF relays because of its simplicity. In the existing works with AF relay, the authors have not considered the direct link signal to noise ratio (SNR) in their problem formulation. Hence, given the power constraint, for some sources, when the direct link SNR is better than the relayed link one, performance of the solution deviates from the optimal one. In this work, we have solved the same problem, however considering the direct link SNR in the problem formulation. Because of incorporating the direct link SNR, the resultant problem is non-convex. We have used geometric programming (GP)-based method to solve this problem. Furthermore, having noticed the high computational complexity of the optimal solution, we also have proposed two suboptimal solutions that take greedy and fair nature of the sources into account.

- **Contribution 3:** In the previous work, all nodes in the network are altruistic. They work selflessly towards maximizing the network capacity. However, in the real world, the user nodes are not that selfless, they want to maximize their own interest while sacrificing their resource. In order to model such selfish behavior of the nodes, we have proposed a game theoretical solution of this problem in this work. Although there are some existing game theoretical solutions for the relay power allocation, there is no complete joint source and relay power allocation solution for this problem. By the help of two Stackelberg games, our proposed game theoretical solution is partially distributed. Finally, through simulation, we have proved that the proposed game theoretical solution achieves comparable performance with the centralized optimal solution.

- **Contribution 4:** Fourth, we have proposed a basic uplink scheduling and resource allocation scheme for LTE heterogeneous traffic systems considering all standard specific constraints and individual user QoS demand. As mentioned in Chapter 2 that there were many works done on the downlink scheduling for homogeneous and heterogeneous traffic networks. Compared to that, not much works have been done on the uplink scheduling for heterogeneous traffic networks. A few works which dealt the uplink scheduling for QoS-aware LTE networks, have not considered detailed standard specific constraints in their formulation. However, in order to measure the system performance precisely, it is required to take all constraints into account. In order to capture the QoS criteria of different traffic, we have adopted a utility function which is already used for the downlink operation of code division multiple access (CDMA)-based systems. Furthermore, to facilitate the interest of the service providers, we consider an additional concept, opportunity cost function which is constructed based on the granular resource utilization. Dual decomposition method has been used in order to solve the formulated problem. Having noticed the high computational complexity of the optimal solution, we have given a suboptimal algorithm with relatively lower complexity. Through extensive simulation, we have justified the efficacy and effectiveness of our scheme comparing with other existing solutions of LTE systems.

1.2 Organization of the Thesis

In Chapter 2, we briefly summarize the background and existing works in order to justify the contributions of this thesis. In Chapter 3, we propose an uplink scheduling and resource allocation scheme for DF relay (with FD functionality) based systems (contribution 1). We propose the centralized solution for joint source and relay power allocation of an AF-relay-based system (contribution 2) in Chapter 4. In Chapter 5, we provide the game theoretical solution of the problem in Chapter 4 (contribution

3). We present an uplink scheduling and resource allocation scheme for a system carrying heterogeneous traffic (contribution 4) in Chapter 6. Finally, in Chapter 7, we summarize the thesis with some directions for future research.

Chapter 2

Background and Related Work

In this chapter, we provide the background materials to support the contribution of this thesis. As mentioned in the previous chapter, the demand for the bandwidth in cellular networks is continuously increasing due to the varying traffic with different QoS requirements. To cope with this increasing demand, we have seen unprecedented levels of improvement in the physical layer. However, this improvement is not sufficient to sustain the satisfactory performance of new applications. Among many possible ways to improve the performance of cellular networks, smart implementation of the radio resource management layer is one given the limited amount of resource. As this thesis is linked to the radio resource management of LTE-Advanced systems, we would like to give an overview of the evolution of such systems and their radio resource management policies. Scheduling and resource allocation is one of the tasks in this layer, and the objective of this thesis is to prove that optimal implementation of this task can enhance the system performance. Although this task is equally important for both the downlink and uplink operations, in this thesis, we will only stick to the uplink scheduling and resource allocation to fill the gap in the literature. Furthermore, LTE-Advanced systems recommend several new techniques to improve the system performance. Relaying is one of the mechanisms. Since two problems of this thesis are analogous to the relays and one problem is pertinent to heterogeneous traffic, we will briefly summarize each of the concepts in separate section.

The rest of the chapter is organized as follows. Section 2.1 briefly summarizes the evolution of LTE-Advanced systems and their radio resource management policies. Moreover, we describe the channel structure of such systems, and define the uplink scheduling and resource allocation in this section. Section 2.2 and Section 2.3 provide the brief overview of relay technologies and emerging heterogeneous traffic of LTE-Advanced systems, respectively. Finally, Section 2.4 summarizes this chapter.

2.1 Evolution of LTE-Advanced Systems

The ever increasing demands of users have triggered researchers and industries to come up with a comprehensive manifestation of the fourth generation (4G) mobile communication system. Looking past, wireless access technologies have followed different evolutionary paths aimed at the unified target: performance and efficiency in the high mobility environment. The first generation (1G) has fulfilled the basic mobile voice, while the second generation (2G) has introduced capacity and coverage. This is followed by the third generation (3G), which has quest for data at higher speeds to open the gates for truly mobile broadband experience, which is further realized by the 4G. LTE-Advanced is considered as one of the 4G technologies. The 4G technologies support advanced mobile (low to high) services for high data rate applications while satisfying the service demands of multi-user heterogeneous traffic environments. In order to achieve optimal operations of the 4G technologies, intelligent implementation of the radio resource management layer is required.

Radio resource can be both time and physical spectrum. All problems in this thesis consider spectrum as radio resource. The limitation in radio spectrum comes from the fact that the spectrum is finite and it is not free. The more wireless applications and technologies are used, the more bandwidth is required. Moreover, physical spectrum alone does not mean bandwidth. Wireless devices need power in order to use the spectrum for physical data transmission. However, the continuous growth in wireless data traffic results in the increase of energy consumed by wireless networks, which leads to undesirable increase in carbon dioxide (CO_2) emission. CO_2 is considered as the chief greenhouse gas that are resulted from wireless networks and other human activities, and causes the global warming and climate changes. Despite there are serious efforts to reduce the amount of CO_2 emission at the base stations and mobile subscribers, efficient power management on the limited spectrum is urgently required.

Although there are several aspects of radio resource management at different levels of the network for achieving different services, this thesis is dedicated to the scheduling of spectrum and power management. Scheduling in LTE like systems can be done at three different levels: admission, class, and packet level. Admission level scheduling is responsible for accepting or rejecting new user connections. It aims at satisfying the long term goal of users by maximizing the number of admitted connections while maintaining the satisfaction of ongoing ones. Class level scheduling deals with the aggregate demand of admitted users. It determines the number of transmission time frames that each class needs in order to maintain the satisfaction of its admitted users. Once the time frames are provisioned between different classes, packet level scheduling is utilized in order to determine which of the user packets are transmitted in a single frame.

This thesis deals with the packet level scheduling. Such schedulers may operate for both the downlink and uplink transmissions. Centralized controller is mainly responsible for any of these scheduling operations. However, for the uplink scheduling, individual mobile nodes may make some autonomous decisions. All problems we consider in this thesis are for the uplink spectrum scheduling and power allocation.

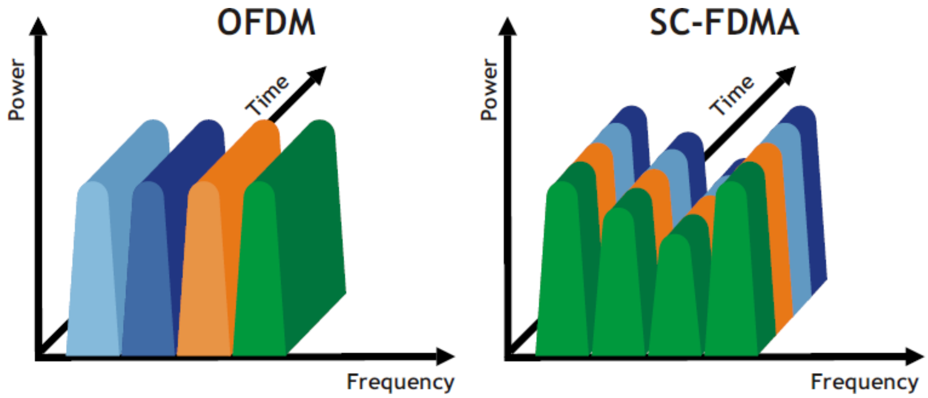


Figure 2.1: In OFDM, each frequency component carries unique information. In SC-FDMA, the information is spread across multiple subcarriers [3].

The following two subsections describe the uplink spectrum access mechanism, and the definition of uplink scheduling and resource allocation, respectively.

2.1.1 Uplink Spectrum Access Mechanism

LTE and LTE-Advanced take advantage of orthogonal frequency division multiple access (OFDMA), a multi-carrier scheme that allocates radio resources to multiple users. OFDMA uses OFDM access technique. SC-FDMA is recommended as the 3GPP LTE uplink transmission technique. It is a modified form of OFDMA, and has similar performance and almost the same overall complexity as OFDMA has. Similar to OFDM, SC-FDMA also multiplexes subcarriers, but it transmits on the subcarriers in sequence not in parallel which is the case in OFDM. This prevents power fluctuations in SC-FDMA signals, i.e., low peak to average power ratio (PAPR) [5]. Figure 2.1 shows the difference between OFDM and SC-FDMA.

As depicted in Figure 2.2, one OFDM/SC-FDMA frame constitutes 20 slots, each being 0.5 ms long. Each slot is called a subframe or TTI.

The structure of one slot can be more clearly understood by looking at the resource

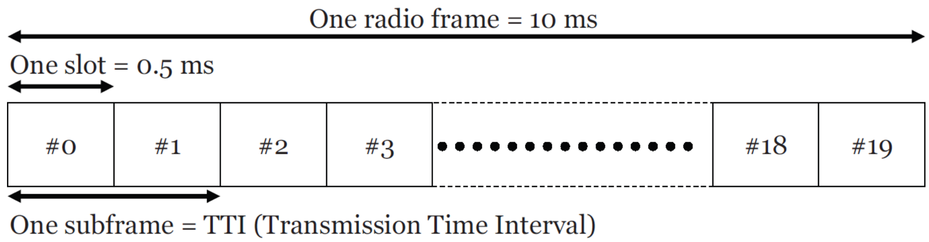


Figure 2.2: LTE frame structure [2].

grid structure in Figure 1.2 of Chapter 1. The transmit signal in each slot is described by a resource grid with $N_{RB}N_{sc}^{RB}$ subcarriers and N_{symb} OFDM/SC-FDMA symbols. The number of subcarriers for each RB N_{sc}^{RB} is standardized as 12 for the LTE uplink. N_{RB} depends on the uplink transmission bandwidth determined for the cell, but should always be between 6 and 110. These numbers correspond to the smallest and largest uplink bandwidth [6], respectively. When time domain is considered, the number of OFDM/SC-FDMA symbols in each slot is 7 for the normal cyclic prefix. However, when the long cyclic prefix is used, this number decreases to 6. From the scheduling point of view at the radio resource management layer, RB is considered as the minimal resource unit to access although each RB can be broken down into resource elements.

2.1.2 Uplink Scheduling and Resource Allocation

The definition of basic scheduling and resource allocation is the selection of a set of users/relays, determining RB-user/relay mapping, and setting of transmit power on each RB. Because of interference, more than 1 user cannot be assigned to the same RB. Moreover, for each node in the network no matter it is base station or mobile user or relay, there is always a maximum limit on the amount of power available for the purpose of transmission. Depending on the type of transmission, the number of maximum power constraints varies as well. For example, for the downlink operation, there is always one power constraint in the scheduling problem no matter the number of users/relays in the system. On the other hand, for the uplink operation, the number of power constraints proportionally increases with the number of users/relays in the system. Because of the growing number of power constraints, solution of the uplink scheduling problem is more computationally intensive compared to the downlink one.

The allocation of RBs to each user is an important issue which has an influence on the system performance of the LTE uplink data transmission. Although SC-FDMA is the recommended technique for LTE uplink data transmission, OFDMA is often used in the literature to show the maximum possible multi-user, multi-channel diversity for OFDM-based systems. Chapter 3 of this thesis assumes OFDMA as the uplink transmission technique. Taking OFDMA into consideration, some of the uplink scheduling and resource allocation works from different aspects, such as channel model, spatial multiplexing, fairness, etc. are [7, 8]. SC-FDMA has two types of RB mapping [9]: Localized FDMA (LFDMA) and Interleaved FDMA (IFDMA). In LFDMA, the scheduler assigns consecutive RBs to convey information from a particular user. In IFDMA, users are assigned RBs that are distributed over the entire frequency band in order to avoid allocating adjacent RBs that are simultaneously in deep fade. Through LFDMA, full multi-user, multi-channel diversity is not achievable due to the contiguity constraint of RBs. Unlike OFDMA, not much works have been conducted on the uplink scheduling for LFDMA technique. Some existing works in this context are [10, 11, 12, 13]. Even lesser works have been done for the uplink scheduling and resource allocation considering IFDMA as the transmission technique.

2.2 Relays in LTE-Advanced Systems

OFDM and OFDMA technologies are well-known for mitigating frequency selective fading and inter-symbol interference which results in increased performance. Even under the deployment of these technologies, nodes at the edge of a cell are often unable to achieve desired performance because of high path loss and shadow fading. In order to tackle this problem, 3G/4G technologies have incorporated relaying concept which is proven to alleviate the dead spots or enhance the coverage for such nodes and improve the network capacity.

Several relay strategy techniques have been proposed in the literature, such as AF and DF (Other less common scheme, such as compress-and-forward (CF)) as forwarding schemes used by the relays [14, 15, 16]. DF strategy is a layer-2 based scheme, where relays act as bridges or routers for the traffic flowing from user terminals to the base station. Through DF, the relay node retransmits just the replica of source node's transmitted signal. Whereas AF means, the relay nodes amplify source node's transmitted signal and then retransmit towards the destination node. The relay operation can further be classified into FD and half duplex (HD) modes. In FD mode, relay node can transmit and receive simultaneously at the same frequency band in one frame, but needs to endure self interference due to the signal leakage between relay output and input. In the HD mode, the relay is restricted to transmit and receive over orthogonal (time division and frequency division) channels for avoiding self interference on it. Previous works mainly concentrated on the HD mode, which enjoys much lower implementation complexity than the FD mode. However, FD is a more challenging and attractive mode which can provide better performance than the HD mode if self interference can be reduced efficiently. The problem in Chapter 3 considers that the system is equipped with DF relays of FD mode, whereas Chapters

4 and 5 assume that the system has an AF relay with FD mode.

Over the last few years, numerous works have been conducted on the basic downlink scheduling for OFDM-based DF relay enhanced cellular networks. In order to maximize the system capacity, some heuristics for subcarrier and power allocation are given in [17], [18] and one detailed thorough work with required proofs is [19]. For minimizing power, there are some contemporary works, e.g., [20]. Once people have got an idea about optimal resource allocation to maximize the overall rate, they started looking at the problem from fairness point of view. For assuring fairness, some works are [21, 22]. Consequently, for ensuring QoS of users, one work is [23]. All these works are based on the assumption that routing information of relays is fixed. Joint intra-cell routing and queue-aware downlink scheduling solutions have been proposed in [24], [25]. [24] is for HD relayed systems, whereas [25] is for FD relayed systems. Taking inter-cell interference into account, optimal resource allocation solution in multi-cell scenario is given in [26]. Furthermore, [27] and [28] have incorporated spatial reuse factor in their resource allocation formulation.

Under the deployment of fixed DF relays, uplink capacity analysis for LTE or OFDMA-based systems has appeared in some recent papers. Reference [29] has studied the uplink capacity of relay enhanced 802.16j networks that have OFDM-based physical layer. Resources they have considered are transmission duration of relays, mobile nodes and their assigned power. Optimizing the subframe duration and power, another work is [30]. Unlike [29], they co-schedule macro cell users and relays at the same subframe, and then optimize their transmission power in order to mitigate inter-cell interference. Subsequent another subframe sharing work in uplink systems is [31]. In [32], network coding has been used for the purpose of resource allocation in OFDMA-based relay enabled systems. Relay node applies network coding between the uplink and downlink transmissions in order to attain its objective. Similar coding-aware another work is [33], however their attained resource allocation is proportionally fair. Despite all these sophisticated advancement on the uplink resource allocation of LTE like systems, basic scheduler (finding RB-user mapping, and their power assignments) is still not studied that much. Chapter 3 proposes a basic uplink scheduler for DF relay enhanced LTE systems. After the publication of our work, a few uplink resource allocation schemes appeared in the literature recently, such as [34, 35, 36]. Unlike us, [34, 35] proposed the resource allocation schemes on the assumption that the physical layer is based on SC-FDMA technique. Whereas, considering OFDMA-based physical layer, the resource allocation scheme of [36] optimizes the routes of all user nodes.

As the deployment cost of relay is high, there might be a few number of relay nodes in the cell which can help the edge nodes to transmit their data. From this perspective, one of the key problems in a relay equipped node is to make decision which edge nodes to be helped and how much power need to be disseminated among them in order to maximize the system throughput. There are some existing works in this context for a DF relay are [37, 38]. However, DF relays decode, re-modulate and retransmit the received signal, and hence incur high computational complexity. Therefore, researchers often recommend to use AF relays because of its simplicity. Power optimization of such networks with AF relays have extensively been studied. At the beginning of such exploration, people focused on a simple network with one source and one relay [39]. And, then, relay power allocation under a single-source multi-relay network has been studied varying different optimization criterion, i.e., outage probability minimization or sum relay power minimization under sources' SNR or outage probability constraints [40], [41], [42].

Performance optimization of multi-source single-relay network has also been given attention from different angles, such as interference cancellation schemes are proposed in [43] and [44]. In [45], network decoding is applied to combat interference among users. At the same time, joint multi-source and multi-relay power optimization has appeared in a few papers [46], [47] under different optimization criterion, i.e., sum rate maximization, sum power minimization, etc. Although in [46], each source is essentially served by only one relay, [47] made performance improvement by assisting single source with the help of multiple relays. The drawback of their approach is, they have totally ignored SNR due to the direct link transmission in their optimization formulation. Their approaches work very well when the direct link channel quality is worse than the relayed link one. However, the performance starts to deviate from the optimal one when the direct link's quality is better than the relayed link one. Having noticed this, we put the direct link SNR in the original formulation, and our work in Chapter 4 is to allocate power among the sources and relay based on that formulation.

In the solution of the problem in Chapter 4, all entities in the network are altruistic, work selflessly towards maximizing the network capacity. However, in the real world, nodes especially the users are not that selfless. They want to maximize their own revenue or utility while sacrificing their resource. In order to model such selfish behavior of the user nodes, game theory is a natural and flexible tool. Moreover, necessity of the game theoretical solution for this problem can be explained in other way. In the altruistic solution of the problem, the users with better channel quality are given more preference since these users contribute more towards the maximization of network capacity. However, this makes those users more valuable compared to other users with worse channel condition, which is essentially not a fair attitude.

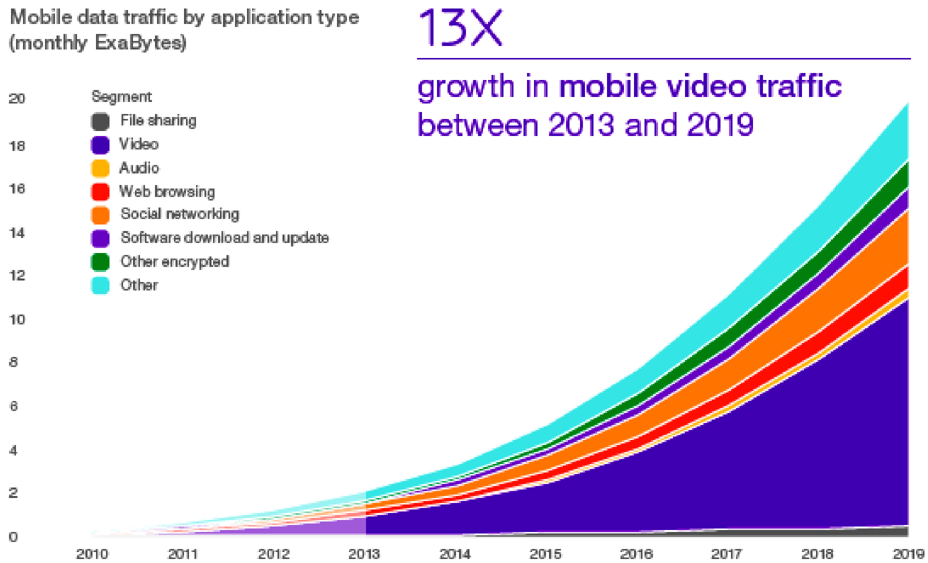


Figure 2.3: The increasing trend of emerging heterogeneous traffic (Ericsson consumer lab) [4].

Game theory allows all users to compete for their interests, i.e., utilities. The winning situation of the game is decided by their actions. In this way, all users are treated in the fair manner. Chapter 5 provides the game theoretical solution of the problem in Chapter 4. With the publication of our work, a few contemporary game theoretical power allocation solutions appeared for slightly different types of relay equipped networks in [48, 49, 50] recently.

2.3 Heterogeneous Traffic in LTE-Advanced Systems

With the rapid advancement of LTE, LTE-Advanced, and other mobile broadband networks, the number of smartphone subscribers have increased throughout the world. Smart phone devices host several applications, including games, video, file sharing, music streaming, etc., which generate heterogeneous traffic in the network. Moreover, due to the participation of large volume of users in the social networking sites, various types of traffic are continuously increasing. In future, we can anticipate more advanced applications which require more resources than the current applications do. Figure 2.3 shows the increasing proportion of data due to different applications in the network over the years. However, to keep the demands of all these applications intact with the growing number of subscribers, it is very essential to design an efficient scheduler which can satisfy the conflicting requirements of different traffic.

QoS ensured resource allocation for OFDM-based networks especially LTE systems has appeared in some survey papers [51], [52]. For homogeneous traffic networks, the downlink scheduling and resource allocation problem has extensively been studied. For example, for delay sensitive traffic, some works are [53, 54], whereas for the traffic with data rate requirement, one work is [55]. Even specifically for VoIP, there are some works, e.g., [56].

The scheduling problem in heterogeneous traffic networks is more challenging than in the homogeneous ones because of the conflicting requirements of different traffic. There are three design issues that need to be considered while formulating the problem of an ideal scheduler for heterogeneous traffic networks. Wireless users experience varying channel quality condition time to time due to the stochastic fading effects, and hence their achievable data rates are affected. The scheduler should choose the user with good channel quality condition in order to maximize the system throughput. However, if the user with better channel quality condition is always selected, the users with worse channel condition may starve. This issue is known as fairness and another very important factor to keep into consideration. Third, different applications have different QoS requirements, and the utility function of the scheduler should be defined in such a way that it has the ability to capture those metrics simultaneously and effectively.

Similar to homogeneous traffic networks, the downlink scheduling and resource allocation has extensively been studied for heterogeneous traffic environments. These works mainly adopted three categories of designs. First, some works gave strict priority to high priority traffic compared to lower priority ones, such as [57, 58]. Second, the authors in [59] have addressed mixed traffic (delay sensitive, guaranteed bit rate, best effort) by designing the scheduler on the time and frequency domains. The priority sets are populated based on the QoS class indicator (QCI) of each data flow and classified as guaranteed bit rate (GBR) and non-GBR. Third, different utility functions are used for different types of traffic (traffic with data rate requirement, traffic with delay constraint) [60, 61]. Besides these, there are some general mechanisms for handling mixed traffic. For example, the authors in [54] have proposed a low complexity RB allocation algorithm using LOG/EXP, earliest deadline first (EDF) rules, and tested their scheme with mixed streaming and live video traffic.

On the other hand, there are very few works for the uplink scheduling with QoS assurance. These works are mainly for homogeneous delay sensitive traffic [62, 63]. One recent work on energy efficient QoS assured scheduler is [64], where QoS is considered as user's instantaneous rate. For heterogeneous traffic environments, recently, [65] has proposed a scheduler which is based on the time and frequency domain scheduler [59]. One drawback of this work is, they did not consider traffic class in their design. Moreover, they evaluated the performance of this scheduler in their customized simulation scenario, which is not practical. Having noticed the drawback of [65] and in order to design a utility based scheduler for heterogeneous traffic environments, Chapter 6 of this thesis proposes an uplink scheduling and resource allocation scheme of LTE systems conveying heterogeneous traffic. Once our scheduling scheme is off the shelf, a few works came forward consequently for heterogeneous traffic networks, such as [66, 67, 68]. In [66], the authors proposed the scheduling scheme considering rate constraint of each user, and evaluated the performance of their scheme in a limited simulation scenario. On the other hand, [67, 68] are for the downlink scheduling.

2.4 Summary of the Chapter

In order to strengthen the contributions of this thesis, in this chapter, we have briefly compiled the evolution of LTE-Advanced systems and their radio resource management policies. Furthermore, we have revised the relay technologies available in the standard of such systems as well as the increasing trend of emerging heterogeneous traffic that is expected to be met by these systems.

Chapter 3

Uplink Scheduling and Resource Allocation for DF Relayed Systems

3.1 Introduction

Uplink scheduling and resource allocation for OFDMA-based systems is a challenging problem because of individual nodes' power constraints and the discrete nature of resource assignment. Emerging 3G/4G cellular network technologies, i.e., LTE, LTE-Advanced are based on OFDMA-based air interface where granular resource to access is named as RB. Scheduling and resource allocation task in a typical LTE system determines the set of users, assignment of RBs to these users and setting of transmit power for each RB. In addition to providing high data rate, OFDMA is well-known for mitigating frequency selective fading and inter-symbol interference. Even under the deployment of OFDMA technologies, far distant nodes of a cell are unable to achieve desired performance because of higher path loss. In order to tackle this problem, 3G/4G technologies have incorporated relaying concept which is proven to enhance the coverage for such nodes and improve the network capacity. In this work, we have considered a network enhanced with fixed digital DF relays deployed by service providers in strategic locations. Scheduling problem appears to be more challenging under such enhancement of networks. Moreover, the blind employment of relays in the network brings a few more key questions. If service providers deploy more and more relays, those may consume system resource while starving regular user nodes which is not an expected characteristic of an ideal resource scheduler. Under the blind deployment of a large number of relays, an ideal scheduler should be able to determine which relays are useful and which are not contributing towards the overall system performance.

In this chapter, we have formulated the uplink scheduling and resource allocation problem of FD relay aided LTE networks as a convex optimization problem. Because of the LTE standard, resource constraint and deployed relays, the resultant problem has three types of constraints. We have used dual decomposition method in order to solve this problem. While exploring the solution structure, we have found, for the given Lagrangian factors due to a relay constraint, the authors in [7] have given many suboptimal algorithms in order to obtain optimization variables, i.e., RB assignments, power allocated on them. We have used their suboptimal algorithms in order to find the principle optimization variables. Revealing the guiding principle of the optimal Lagrange multipliers due to the relay constraints, we have proposed an iterative procedure in order to obtain the suboptimal value of these multipliers, which essentially optimizes resource allocation across the network. The iterative procedure has the ability to deactivate relays if they hamper the performance of other independent nodes in the system. The nodes connected to a deactivated relay act as independent nodes. Because of the fast fading, the iterative procedure may not be suitable to be deployed as a scheduler, and hence a suboptimal algorithm has also been proposed for RBs and their power assignments with the help of one heuristic proposed by [7]. Afterward, we have shown that if the proposed problem is formulated in the game theoretical way, the resultant solutions ensure fairness across the user nodes. An extensive simulation has been conducted in order to justify that deploying some low cost relays, we can essentially improve the system performance in terms of overall throughput and occasionally fairness.

The remainder of this chapter is organized as follows. A brief overview of literature relevant to our work is given in Section 3.2. In Section 3.3, we explain the terminology used for the uplink scheduling of relay aided networks, and formulate the optimization problem. Solution approach of the formulated problem is presented in Section 3.4. Based on the solution approach, our designed algorithms are given in Section 3.5. We provide the simulation results in Section 3.6, and finally Section 3.7 concludes this chapter.

3.2 Related Work

Over the last few years, numerous works have been conducted on scheduling and resource allocation of OFDM-based relay enhanced cellular networks. In [17], Huang et. al. proposed a centralized heuristic solution in order to perform subcarrier and power allocation resulting in the capacity maximization of downlink systems. Similar other works are [69], [70] and [71]. Minimizing total power subject to the rate constraint in such systems first appeared in [72]. In [20], the authors have studied similar problems proposed in [17], [72], however they have improved the results by finding optimal frame duration for the base station and relays. Most recent sum rate maximized resource allocation for downlink relayed systems is proposed by [18] subject to the sum sources and relays power constraint. Optimal resource allocation while ensuring max-min fairness has been proposed in [73], [21]. Whereas, for ensuring proportional fairness, some studies are [22], [74]. QoS-based subchannel and power allocation was proposed by [23] in order to ensure QoS requirements of both direct and relayed users.

Joint intracell routing and queue aware downlink scheduling solutions have been proposed in [24], [25]. [24] considers the relays are of quasi FD mode, whereas [25] shows the performance improvement compared to the other one by adopting the relays with HD mode. Furthermore, [25] has conducted performance evaluation on asymmetric traffic, whereas the former one performed that on symmetric traffic. [75] has presented a scheduling and resource allocation scheme over broadcast block fading channels in order to maximize long term average achievable rate region of the users. Instead of single cell, [26] proposed a resource allocation scheme for OFDM-based DF relayed systems taking multi-cell interference and heterogeneous user rate requirements into account. In order to achieve stable data rate for multimedia users, the authors in [76] have proposed a downlink scheduling technique for MIMO OFDM-based macro cells where there are some fixed relays. Besides proposing typical scheduling algorithms for relay enhanced networks, there are some works done in the last couple of years targeting on further performance improvement by exploiting some environmental or inherent factors. While exploiting the multi-user and multi-channel diversity of OFDM-based systems, [77], [27] and [28] have incorporated another factor, spatial reuse while developing their scheduling algorithms.

All works discussed above are for the downlink scheduling on more or less HD relayed OFDM-based systems. To date, not many works have been conducted on the uplink resource allocation for OFDMA-based relay enhanced cellular networks. Given fixed routing information, [78] has proposed a subcarrier and power allocation scheme for OFDMA-based wireless mesh networks with the objective of maximizing Nash bargaining fairness criterion. Unlike us, their scheduling scheme is partially distributed, each mesh router is independently responsible to allocate number of subcarriers to its mesh clients, later each mesh client solves a mixed integer programming problem for detailed resource allocation. Reference [29] has studied the uplink capacity of relay enhanced 802.16j networks with OFDMA-based physical layer. Resources they have considered are transmission duration of relays, mobile nodes and their assigned power. Unlike us, one assumption of their work is, the relay or mobile node gets hold of all RB like subcarriers of its allotted transmission duration. In [32], network coding has been used for the purpose of resource allocation in OFDMA-based relay enabled systems. Relay node applies network coding between the uplink and downlink transmissions in order to attain its objective. Another sum rate maximization work is [19] where the main assumption is, users are served by multiple DF relays instead of one. Taking individual power constraint for users and relays, they proposed a few suboptimal algorithms upon the non-convexity proof of the original problem. Similar to our work, uplink subchannel and power allocation has appeared previously in one paper [79] for OFDMA-based systems. However, that paper assumed that the relays are with HD functionality instead of FD.

3.3 System Model and Problem Formulation

3.3.1 Problem Statement

We consider a typical LTE cellular network, in which N_T users transmit to the same base station. There are some nodes across the cell edge, which suffer from higher shadowing effect due to the long distance from the base station. In order to remedy this shadowing effect, a set of N_K relays are deployed at reasonably fair distance from the base station. These relay nodes help transmit protocol data units (PDUs) from the base station to the edge nodes, and vice versa. Similar to the terminal nodes, their physical layer is based on OFDMA subcarriers combined in time division multiple access (TDMA) RBs. Since relay nodes serve a number of terminal nodes, transmission power of the relays is considerably larger than that of typical user nodes. In addition, any relay is assumed to have the ability to receive and transmit concurrently using orthogonal RBs. In order to mitigate the self interference on the transmit and receipt RBs, the relays are equipped with separate antennas for two distinct operations. Our goal in this work is to allocate RBs across the user terminals and relays to maximize their aggregate utility.

3.3.2 Motivation Behind Deploying Relays

Our sample network is presumed to be deployed in a rural environment. The channel gain for a particular node t over RB j in such an environment can be represented by the following equation [8][1]

$$H_{t,j,dB} = (-\kappa - \lambda \log_{10} d_t) - \xi_{t,j} + 10\log_{10} F_{t,j}. \tag{3.1}$$

In Equation 3.1, the first term κ captures propagation loss, d_t is the distance in km from mobile terminal t to the base station. λ is the path loss exponent, which is set to a value of 3.76. The second term, $\xi_{t,j}$ captures the log-normal shadowing effect for the reference distance 1 km. $F_{t,j}$ corresponds to the Rayleigh fading effect with a parameter a such that $\mathrm{E}[a^2] = 1$. In the subsequent discussions, we will see that there is a power constraint on each node and power is equally subdivided among the allocated RBs of a user according to the proposed suboptimal algorithms in [7]. Under this condition and because of the parameter assigned for the Rayleigh fading effect on each RB, each node cannot get as many RBs as available since after a few

[1] Instead of rural environments, if some other environments, such as urban, semi-urban would be considered, the channel model would be different. Hence, the resultant outcome of our proposed scheduling algorithms would be different.

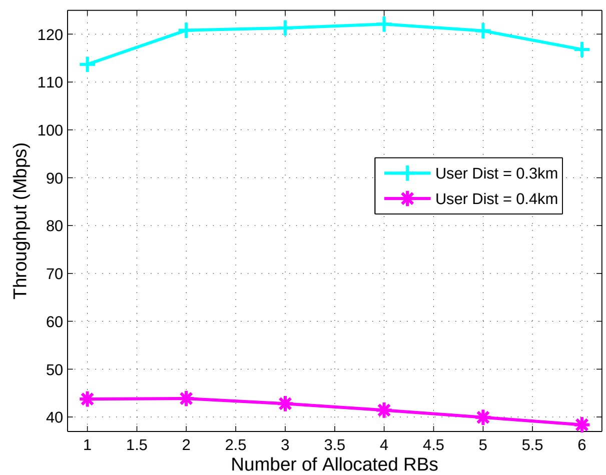


Figure 3.1: Individual user's throughput w.r.t. its allocated RBs.

allocations, its achieved rate deteriorates. Figure 3.1 justifies this argument which depicts the rate of 2 users at different distance from the base station with respect to the number of allocated RBs. We see, after allocating the 4th RB, the rate for the user with distance 0.3 km gets degraded and it keeps degraded. Similar situation is for the other user, however it happens after allocating different number of RBs. Therefore, if the load is low in the network, we will see later that many RBs remain unallocated. Observing these unused RBs and worse condition of the edge nodes, we believe, deploying some relays will enhance the performance of the edge nodes as well as the overall system while consuming the redundant RBs. Moreover, we have seen the results under different deployment scenarios that even if a relay takes users' resource away, achieved throughput using that relay is larger than that without using it.

3.3.3 Problem Formulation

We denote a set $\mathbf{T} = \{1, 2, \cdots, N_T\}$ for the terminal nodes, a set $\mathbf{K} = \{1, 2, \cdots, N_K\}$ for the relay nodes. Total number of RBs available in each scheduling epoch is N_{RB} and the corresponding set for this is $\mathbf{R} = \{1, 2, \cdots, N_{RB}\}$. Although each RB is equivalent to 12 subcarriers, in the LTE standard, granular resource access by user is RB. $p_{t,j}$ and $p_{k,j}$ are the instantaneous transmission power of user terminal t and relay node k, respectively on RB j. For the uplink transmission, each node has constraint on the maximum power it can use at each scheduling epoch, and thus P_t and P_k are the maximum power for the terminal and relay nodes, respectively. We denote a set $\mathbf{P}$ meeting the requirement of these constraints. These can be represented by

$$\sum_{j}^{N_{RB}} p_{t|k,j} \ \leq \ P_{t|k}, \ \forall t \in \mathbf{T}, \ \forall k \in \mathbf{K}. \tag{3.2}$$

Let $x_{t,j}$ denote the fraction of RB j allocated to node t, where the total allocation across all nodes should be no larger than 1, i.e.,

$$\sum_{t=1}^{N_K+N_T} x_{t,j} \ \leq \ 1, \forall j \in \mathbf{R}, \tag{3.3}$$

where $\mathbf{x}$ denotes the set satisfying the constraint in Equation 3.3. According to the LTE standard, there is a constraint that each RB cannot be assigned to more than one node, i.e., $x_{t,j} \in \{0, 1\}$. In the subsequent discussions, we will see how we have dealt this constraint.

The relay nodes are configured in such a way that they have more capability compared to the ordinary terminal nodes. Besides, one relay serves a number of terminal nodes, and therefore the total rate achieved by the terminal nodes cannot be more than that of the corresponding relay, and vice versa. In order to ensure proper utilization of achieved rate, we have used approximation sign in the constraint. Less than sign would work as well. However, it would cause wastage of consumed resource by the relays. Let there are N_{T_k} number of terminal nodes connected to the kth relay, and the corresponding set is $\mathbf{T}_k$. Denote the instantaneous rate of the kth relay is r_k, whereas that of the tth user connected to the kth relay is r_{tk}. According to this constraint

$$\sum_{t\in\mathbf{T}_k} r_{tk} \ \approx \ r_k, \quad \forall k\in K. \tag{3.4}$$

The utility function is defined as service satisfaction of each individual node in terms of throughput, delay or other QoS criteria. The optimization problem is to allocate a set of RBs $I_{RB,k}$ or $I_{RB,t}$ to each relay or terminal node such that the sum utility of all nodes is maximized. This can be formulated in the following way

$$\textbf{maximize} \sum_{t=1}^{N_T} U(I_{RB,t}, P_t) + \sum_{k=1}^{N_K} U(I_{RB,k}, P_k). \tag{3.5}$$

In this work, we first define the utility function of a node as its mere rate. Then, for the second step, in [78], we have seen, if the objective function is the product of all nodes' rate (Nash product), the resultant resource allocation is fair across them. It is a well-known result in game theory that the solution to the cooperative bargaining problem maximizes the Nash product [80]. The objective function is further modified by taking logarithmic of it knowing the fact that logarithmic function is strictly concave. Hence, the resulting objective function becomes the sum of logarithmic rate for all nodes which is proven to ensure Nash bargaining fairness. In order to ensure fair resource allocation among all nodes, we have studied the utility function as logarithmic of node's rate. When the utility function is mere rate, the objective function yields

$$\arg\max_{\mathbf{x},\mathbf{P}} \sum_{t\in(\mathbf{T}\cup\mathbf{K})} r_t;\ r_t = \sum_{j\in\mathbf{R}} x_{t,j}\log\left(1+\frac{p_{t,j}e_{t,j}}{x_{t,j}}\right), \tag{3.6}$$

subject to the constraints in Equations 3.2, 3.3 and 3.4. $e_{t,j}$ represents the SNR per unit transmit power for node t on RB j. Typically, $e_{t,j}$ is named as channel gain. Since each RB consists of 12 subcarriers, gain $e_{t,j}$ is the average of all subcarriers on RB j. Moreover, while computing the gain, the physical distance for the terminals connected to a relay is determined assuming the end point as that relay. Whereas, for other nodes including the relays, the end point is the base station or centralized controller.

In this section, we have introduced the sample network architecture, motivation for relay based systems, and then formulate the problem after defining all terminologies to explain the uplink scheduling problem.

3.4 Solution Approach

As the number of relays in the system is N_K, the total number of terminal nodes connected to the relays is $N'_T = \sum_{k\in\mathbf{K}} N_{T_k}$. The rest of the terminal nodes are independent, and they directly send their data to the base station and this number is $N''_T = N_T - N'_T$; the corresponding set is $\mathbf{T}''$. Taking these issues into consideration, the objective function in Equation 3.6 can further be broken down into

$$\arg\max_{\mathbf{x},\mathbf{P}} \chi(\mathbf{x},\mathbf{P}) := \sum_{t\in\mathbf{T}''} r_t + \sum_{k\in\mathbf{K}} r_k + \sum_{k\in\mathbf{K}}\sum_{t\in\mathbf{T}_k} r_{tk}. \tag{3.7}$$

The objective function in Equation 3.7 has no duality gap and we can solve it by considering a dual formulation. Keeping the constraints in Equations 3.2 and 3.3 intact and taking dual variables $\mathbf{\Lambda} = (\Lambda)_{k\in\mathbf{K}}$ for the constraint in Equation 3.4, the resulting Lagrangian becomes

$$\begin{aligned}L(\mathbf{\Lambda}, \mathbf{x}, \mathbf{P}) &:= \sum_{t\in\mathbf{T}''} r_t + \sum_{k\in\mathbf{K}} r_k + \sum_{k\in\mathbf{K}}\sum_{t\in\mathbf{T}_k} r_{tk} \\ &\quad + \sum_{k\in\mathbf{K}} \Lambda_k \left(\sum_{t\in\mathbf{T}_k} r_{tk} - r_k\right) \\ &:= \sum_{t\in\mathbf{T}''} r_t + \sum_{k\in\mathbf{K}} (r_k - \Lambda_k r_k) + \sum_{k\in\mathbf{K}}\sum_{t\in\mathbf{T}_k} (r_{tk} + \Lambda_k r_{tk}).\end{aligned} \tag{3.8}$$

According to the duality theory, the optimal solution of Equation 3.7 can be given by

$$\arg\min_{\mathbf{\Lambda}} \arg\max_{(\mathbf{x},\mathbf{P})} L(\mathbf{\Lambda}, \mathbf{x}, \mathbf{P}). \tag{3.9}$$

The problem in Equation 3.8 has the similar form as the formulated problem without a relay in [7] except the rate function for the relays and the terminals connected to the relays are modified by adding additional terms due to the constraint in Equation 3.4 and its associated dual variables. The way of obtaining optimal solution vector $(\mathbf{x}, \mathbf{P})$ for this problem has been discussed in [7]. Having observed the high computational complexity of the optimal solution, the authors in [7] have proposed a few suboptimal algorithms. We plan to use these suboptimal algorithms (SOA1 and SOA2) in order to obtain resource vectors $(\mathbf{x}, \mathbf{P})$ for the given $\mathbf{\Lambda}$. We have extended both heuristics in order to achieve our desired objective. SOA1 has also two variants. We have used variant 5A since variant 5B uses constant power for each allocated RB which is not pragmatic. In this section, we have used SOA1(5A) in order to obtain $\mathbf{x}$ and $\mathbf{P}$. SOA1 runs N_{RB} number of times, and in the jth iteration, node $t\in\mathbf{T}\cup\mathbf{K}$ computes the following metric for its best RB, denoted by $l_t(j)$

$$g_t(j) = \sum_{j\in\Omega_t(j-1)\cup l_t(j)} \log\left(1+\frac{P_t e_{tj}}{|\Omega_t(j-1)|+1}\right) - \sum_{j\in\Omega_t(j-1)} \log\left(1+\frac{P_t e_{tj}}{|\Omega_t(j-1)|}\right), \tag{3.10}$$

where $\Omega_t(j)$ denotes the set of the tth node's accumulated RBs in the jth iteration. We define the first and second terms of Equation 3.10 as $g_t^a(j)$ and $g_t^b(j)$, respectively. The metric defined in such way is for the throughput maximization scheme, whereas for ensuring Nash bargaining fairness, metric $g_t(j)$ is considered as $\log(g_t^a(j)) - \log(g_t^b(j))$ which has been proved in [8]. The following lemma shows and proves the way this metric has been computed for our relay oriented problem.

Lemma 3.1: For relay $k\in\mathbf{K}$, $g_k(j)$ is computed as $(1-\Lambda_k)g_k^a(j) - (1-\Lambda_k)g_k^b(j)$ and for its descendents $t\in\mathbf{T}_k$, $g_t(j)$ is computed as $(1+\Lambda_k)g_t^a(j) - (1+\Lambda_k)g_t^b(j)$.

Proof : If we compare the uplink scheduling problem of [7] and our Equation 3.8, we notice, the rate function of the kth relay is replaced by $(r_k - \Lambda_k r_k)$ and the rate function of its descendents is replaced by $(r_{tk} - \Lambda_k r_{tk})$, rest of the other nodes' contribution towards the objective function are as same as it is in Equation (UL) of [7]. The selection criterion of OFDM tone in SOA1 (5A)[7], $g_i(n)$ is just the function of rate accompanied with the OFDM tones from sets $K_i(n)$ and $K_i(n) \cup l_i(n)$. We redefine set K as Ω. Since the selection criterion is just a rate function of some set with RBs, it is for the kth relay would be set as $g_k^a(j) - g_k^b(j) - \Lambda_k(g_k^a(j) - g_k^b(j))$, i.e., $(1-\Lambda_k)g_k^a(j) - (1-\Lambda_k)g_k^b(j)$. Similarly, for its descendant nodes ($t\in\mathbf{T}_k$), $g_t(j)$ is defined as $g_t^a(j) - g_t^b(j) - \Lambda_k(g_t^a(j) - g_t^b(j))$, i.e., $(1+\Lambda_k)g_t^a(j) - (1+\Lambda_k)g_t^b(j)$.

Once we obtain the suboptimal $\mathbf{x}$ and $\mathbf{P}$, and we substitute in Equation 3.8, the right hand side of the equation becomes the function of $\mathbf{\Lambda}$, i.e., $L(\mathbf{\Lambda})$. The solution to Equation 3.7 is given by minimizing $L(\mathbf{\Lambda})$ over $\mathbf{\Lambda} >= 0$. Moreover, $L(\mathbf{\Lambda})$ is discontinuous and non-differentiable with respect to $\mathbf{\Lambda}$. Hence, we use subgradient based search approach in order to obtain optimal $\mathbf{\Lambda}$. In each iteration of the search method, for the given $\mathbf{\Lambda}$, the suboptimal algorithms [7] find resource vectors $(\mathbf{x}, \mathbf{P})$. Iterations continue until the optimal $\mathbf{\Lambda}$ is obtained. Each step of the search method updates Λ_k in the following way

$$\Lambda_k(u+1) = \left|\Lambda_k(u) + \kappa_k \left|\sum_{t\in\mathbf{T}_k} r_{tk} - r_k\right|\right|, k = 1, \cdots, N_K.$$

The value of $\Lambda_k(u)$ can be both positive or negative because of the equality constraint associated with this dual variable. If we look at Equation 3.8, it is rational to say that the absolute value of Λ_k is in between 0 and 1. This is because, no matter the value of κ_k is positive or negative, if its absolute value becomes greater than 1, either the terminal nodes' ($t\in\mathbf{T}_k$) or the kth relay's individual utility turns out to zero or negative value which is not a desired expectation of this maximization problem. In order to obtain optimal $\mathbf{\Lambda}$, the ideal requirements are the continuous nature of function $L(\mathbf{\Lambda})$ and the optimal value of $L(\mathbf{\Lambda})$ is known. Since the Lagrangian function does not have any of the properties, we have proposed an iterative, suboptimal, adaptive approach in order to find $\mathbf{\Lambda}^*$ which gives more privilege to the independent user nodes. At the first iteration, $\left|\frac{\partial L}{\partial \Lambda_k}\right|$ gets the largest value, the following iterations gradually reduce this term until Λ_k^* is achieved. Given this observation, κ_k is set only at the first iteration, i.e, $\left|\frac{\epsilon}{\sum_{t\in\mathbf{T}_k} r_{tk} - r_k}\right|$, this quantity remains constant in the subsequent iterations. How does ϵ look like, is given in *Observation 3.1*.

Observation 3.1: Because of the discontinuity nature of $L(\mathbf{\Lambda})$ and also from the simulation, we observe, if the value of ϵ is ≤ 0.1, nearly optimal solution is achievable[2]. With large value of ϵ which is > 0.1, the iterative $\mathbf{\Lambda}$ update procedure requires less number of iterations to achieve convergence, however the accuracy deteriorates.

[2]Setting ϵ as ≤ 0.1 provides the reasonable tradeoff between the accuracy and speed.

Therefore, the value of ϵ is the tradeoff between the accuracy and speed.

3.5 Proposed Algorithms

3.5.1 Iterative $\boldsymbol{\Lambda}$ Update Procedure

An iterative algorithm for updating $\boldsymbol{\Lambda}$ using suboptimal algorithm SOA1 [7] is given in *Algorithm 1*. This algorithm is adaptive, if there is a lack of resource or setup of any relay and its terminal nodes is infeasible, dynamically it can disable that relay or can disconnect terminal nodes from the corresponding relay. In this subsection, we will explain how this algorithm works. Before getting into the loop of obtaining resource vectors $(\mathbf{x}^*, \mathbf{P}^*)$, we will discuss different categories of relay placement. The positions of relays and their terminal nodes can be infeasible. Under the infeasible setup of any relay, the inner loop for obtaining the principle resource vectors does not execute. We define the feasibility of a relay as a setup, when (1) there are multiple terminal nodes connected to that relay, the rate obtained by the relay is larger than the sum rate of its terminal nodes, (2) there is only one terminal node connected to that relay, achieved rate by the relay can be lower compared to the terminal one, and vice versa. The rest of the other possible setups are considered as infeasible. Lines $[7-21]$ are for checking the feasibility of each relay and turns them to the feasible one if necessary. The vector $\boldsymbol{\Lambda}$ does not get updated unless all relays in the network yield to the feasible ones. If any relay is found such that its rate is lower than the sum rate of its terminal nodes (attached to it), the algorithm looks for the terminal node with the largest (lowest) rate, disconnects it from that relay, makes it independent and runs the inner loop resource allocation algorithm. This operation for that relay continues unless the rate re-achieved by the relay is larger than the sum rate of its descendent nodes, or there is only one node remaining attached node to that relay. Once all relays are feasible, resource allocation is conducted across all nodes in the network taking vector $\mathbf{\Lambda}$ into account. $\mathbf{\Lambda}$ is updated in each iteration and the inner loop for the resource allocation is re-executed for the updated $\mathbf{\Lambda}$. This procedure continues unless all relays in the network achieve convergence. The convergence for a relay is achieved when its rate is approximately equal to the sum rate of its attached nodes, slightly larger rate of the relay is preferred. While continuing the $\mathbf{\Lambda}$ update procedure, any or some relays can be infeasible. Infeasibility happens when either the rate of a relay or a terminal node becomes 0. However, if a relay is in the rate decrementing process, obtaining 0 rate is considered as convergence for that relay if there is no starved independent node in the network. The reasons behind obtaining 0 rate in each iteration: (1) if the relay is in the process of decrementing its rate and its channel quality in any RB is so good or it has more usable power that it cannot allocate any RB meeting the upper limit of the decremented rate, (2) if the relay is in the process of incrementing its rate and its position is not that good or there is a lack of RB due to the consumption by others, it cannot allocate any RB while satisfying the lower limit of the incremented rate. Similar phenomenons go with the terminal nodes attached to the relays. In either case, if the achieved rate for the relay or any of its terminal nodes appears to 0, that corresponding relay becomes infeasible. If the rate of all terminal nodes attached to any relay becomes 0 or the relay's rate becomes 0 and there is some starved independent node, that relay is forced to be disabled. Whenever such situations happen, the $\mathbf{\Lambda}$ update procedure is reinitialized (Lines $[35-37]$ in the algorithm). In the algorithm, we have defined a flag vector $\mathbf{F}$ for all nodes: 0 for the independent node, $-1/-2/-3$ for the relay when its rate is larger than the sum rate of its terminal nodes, when its rate is lower than the sum rate of its descendants and when it is disabled, respectively. The node with the flag value greater than 0 is for a terminal node attached to any relay and the flag denotes the index of the corresponding relay. Vector $\boldsymbol{\theta}$ keeps the track of feasibility status for all relays.

As we discussed before resource allocation procedure is inside the inner loop, it is almost similar to SOA1 algorithm presented in [7] except the marginal utility $g_{k|t}(j)$ is computed using the formula $(g^a_{k|t}(j) \pm \Lambda_k g^a_{k|t}(j)) - (g^b_{k|t}(j) \pm \Lambda_k g^b_{k|t}(j))$ for the kth relay or the tth terminal node ($t \in \mathbf{T}_k$) attached to it. Instead of computing $g_{k|t}(j)$ using $\boldsymbol{\Lambda}$, we can compute it in a regular manner as presented in [7], but for the relay or terminals nodes, we need to keep another variable which is called virtual rate. The virtual rate is computed in each iteration after resource allocation is performed. In the algorithm, virtual rate vector is defined as $\mathbf{V}$. At the beginning, the virtual rate for all nodes is initialized as ∞. For the independent nodes, over the iterations of the algorithm, it remains unchanged. However, for the kth relay, the virtual rate is $(r_k \mp \Lambda_k r_k)$ and for the $t \in \mathbf{T}_k$th node, it is $(r_t \pm \Lambda_k r_t)$. $\pm$ sign is for the flag value -1 and -2 of the kth relay, respectively. Similar computation goes for the other relays. *Lemma 3.2* proves that resource allocation using virtual rate is approximately equivalent to the original procedure.

Lemma 3.2: Inside the inner loop, if the flag value of the kth relay is -1, its upper bound of rate is $r_k - \Lambda_k r_k$, its $t \in \mathbf{T}_k$th descendent's lower bound of the rate is $r_{tk} - \Lambda_k r_{tk}$. Whereas, when the flag of the kth relay is -2, its upper and tth descendent's lower bound of the rate are $r_k + \Lambda_k r_k$ and $r_{tk} - \Lambda_k r_{tk}$, respectively. The upper bound of the rate for the independent nodes is considered as ∞. Having considered these assumptions, the selection criterion of RB follows the same procedure as in SOA1(5A) [7].

Proof: In the original procedure illustrated in *Lemma 3.1*, we have seen, the RB selection criterion $g_k(j)$ for the kth relay has additional term $\Lambda_k(g_k^a(j) - g_k^b(j))$ subtracted from it. Similarly, for the kth relay's descendent node $t \in \mathbf{T}_k$, $g_t(j)$ has the additional term $\Lambda_k(g_t^a(j) - g_t^b(j))$ added to the original one. It means, no matter for the relay or its adjacent nodes, their instantaneous rate is just subtracted/added by some term which is proportional to Λ_k. Intuitively, instead of computing $g_{k|t}(j)$ using Λ_k, we can use some sort of customized rate to limit their upper or lower bound. While doing resource allocation inside the inner loop, we do not know in advance the instantaneous rate of all nodes, hence we take the rate obtained in the previous iteration and compute their customized rate using $\Lambda_k(\forall k \in \mathbf{K})$ as depicted in *Lemma 3.2.* The resulting resource allocation produces approximately the similar outcome as given by the original procedure. The above arguments are for the relay whose flag is -1. In the similar manner, for the flag value -2, it can be justified that considering the upper bound of virtual rate for the kth relay as $r_k + \Lambda_k r_k$ and its connected $(t \in \mathbf{T}_k)$th node's lower bound of virtual rate as $r_{tk} - \Lambda_k r_{tk}$, if we follow SOA1 [7], the resultant scheduling decision does not make much difference. The RB selection criterion of the independent nodes is not influenced by the dual variables of the relays, and hence their upper bound of rate is ∞.

3.5.2 Low Complexity Suboptimal Algorithm

As explained above, *Algorithm 1* undergoes a number of iterations before all relays in the network achieve convergence. However, in the fast fading environment, the scheduler may need to take quick RB allocation decision because of smaller scheduling interval. Considering this issue, we have presented a suboptimal *Algorithm 3* which performs scheduling decision without considering $\mathbf{\Lambda}$. This algorithm is similar to the original RB allocation algorithm, SOA1 except a terminal node connected to the relays checks feasibility of its current possible RB allocation with respect to its relay which is depicted in the lines $[10-14]$ of *Algorithm 3*. The way of maximizing overall rate and ensuring fairness across the nodes, the criterion of SOA1 works following the similar manner explained in the previous section.

3.5.3 SOA2 Extended Algorithm

We have extended SOA2 [7] as follows. This algorithm has 2 steps: RB number assignment $n_t, t\in\mathbf{T}\cup\mathbf{K}$ and actual RB-node matching determination. In order to achieve the first objective, we need to solve the following problem in Equation 3.11

$$\begin{aligned} &\max_{n_t>=0, n_k>=0} \sum_{t\in\mathbf{T}''} U'_t + \sum_{k\in\mathbf{K}} U'_k + \sum_{k\in\mathbf{K}} \sum_{t\in\mathbf{T}_k} U'_{tk} \\ \mathbf{s.t.}: &\sum_{t\in\mathbf{T}\cup\mathbf{K}} n_t <= N_{RB}, \quad \sum_{t\in\mathbf{T}_k} r'_{tk} \approx r'_k, \quad \forall k\in K, \end{aligned} \tag{3.11}$$

where $r'_t = \sum_{j=1}^{n_t} \log(1 + \frac{P_t}{n_t} e'_{t,j})$. For the throughput maximization scheme, $U'_t = r'_t$. and for ensuring fairness, $U'_t = \log(r'_t)$. $\mathbf{e}'_t$ is the sorted vector of $\mathbf{e}_t$ over all RBs in the descending order. r'_k and r'_{tk} have the similar interpretation as r'_t has. The objective function in Equation 3.11 is a concave problem over a non-integer convex set $\{n_t >= 0, t\in\mathbf{T}\cup\mathbf{K}\}$ and does not have duality gap. Taking the help of virtual throughput discussed in *Algorithm 1* and the dual variables, the resultant Lagrangian becomes

$$L(\mathbf{n}, \lambda) := \sum_{t \in \mathbf{T}'', r'_t <= V_t} U'_t + \sum_{k \in \mathbf{K}, U'_k <= V_k} U'_k$$
$$+ \sum_{k \in \mathbf{K}} \sum_{t \in \mathbf{T}_k, r'_{tk} <= V_{tk}} U'_{tk} - \lambda(\sum_{t \in \mathbf{T} \cup \mathbf{K}} n_t - N_{RB}).$$

For a given λ, the optimal solution of above problem can be obtained by solving the following $N_T + N_K$ sub-problems. We can solve each sub-problem by a bisection search over the range $[0, N_{RB}]$.

$$n_t(\lambda) := \underset{n_t, r'_t <= V_t}{\arg\max} \left(U'_t - \lambda n_t \right), \forall t \in \mathbf{T} \cup \mathbf{K}. \tag{3.12}$$

Let the solution of all sub-problems are $n^*_t(\lambda), t \in \mathbf{T} \cup \mathbf{K}$, and the resultant utilities are $U^*_t(\lambda), t \in \mathbf{T} \cup \mathbf{K}$. Substituting these back to the Lagrangian

$$L(\lambda) := \sum_{t \in \mathbf{T}''} U^*_t + \sum_{k \in \mathbf{K}} U^*_k + \sum_{k \in \mathbf{K}} \sum_{t \in \mathbf{T}_k} U^*_{tk} - \lambda(\sum_{t \in \mathbf{T} \cup \mathbf{K}} n^*_t(\lambda) - N_{RB}),$$

which is a convex function of λ. The optimal value of λ results in the minimum value of $L(\lambda)$ and can further be obtained by another bisection search over the range $[0, max\{\max_{t \in \mathbf{T}} U'_t, \max_{k \in \mathbf{K}} U'_k\}]$. The computational detailed complexity of each procedure is given in [7]. Integer approximation of n^*_t, $t \in \mathbf{T} \cup \mathbf{K}$ is performed as follows. We sort the nodes according to the mantissa of n^*_t, $\mathrm{fr}(n^*_t) = n^*_t - \lfloor n^*_t \rfloor$. It produces a node permutation set α_t, $t \in \mathbf{T} \cup \mathbf{K}$ such that $\mathrm{fr}(n^*_{\alpha_1}) >= \mathrm{fr}(n^*_{\alpha_2}) >= \cdots$. Then, we calculate the number of unallocated RBs, $N^A_{RB} = N - \sum_{t \in \mathbf{T} \cup \mathbf{K}} \lfloor n^*_t \rfloor$. Finally, we adjust the nodes with large mantissas such that all unallocated RBs, N^A_{RB} are allocated, i.e., $n^*_{\alpha_t} = \lfloor n^*_{\alpha_t} \rfloor + 1$.

Once we have RB number assignments $n_t^*, t{\in}\mathbf{T}{\cup}\mathbf{K}$, our next objective is to find out actual node-RB mapping. For this, we have followed exact SOA2-CUM procedure of [7]. For this purpose, they have used *Hungarian Algorithm* which requires virtual node splitting. We denote U_{tj} as the utility of the tth node over RB j, and $\mathbf{U}_t$ as the utility vector over all RBs, $j{\in}\mathbf{R}$. For the throughput maximization and fair schemes, U_{tj} is r_{tj} and $log(r_{tj})$, respectively; r_{tj} is the rate of the tth node on RB j. In the similar manner, if we denote the utility vector for other nodes, we can make matrix $\mathbf{\Gamma} = \left[\mathbf{U}_1^T, \mathbf{U}_2^T, \cdots\right]^T$ with size $(N_T + N_K){\times}N_{RB}$. Next, we need to split each node t into n_t^* virtual nodes by adding $n_t^* - 1$ copies of the row vector to matrix $\mathbf{\Gamma}$. The size of the expanded matrix would be $(N_T + N_K){\times}N'_{RB}$, where $N'_{RB} <= N_{RB}$. Now, if we define the permutation matrix $\mathbf{C}$ with size $(N_T + N_K){\times}N'_{RB}$ and solve problem (19) of [7], we obtain our desired node-RB mapping matrix $\mathbf{C}^*$. If we apply *Hungarian Algorithm* to our problem without any modification, the resultant kth relay's rate might become lower than $\sum_{t\in\mathbf{T}_k} r_{tk}$, which is not feasible. In order to make all relays feasible, we need to follow either of the following steps.

- While solving the problem in Equation 3.11, if we obtain our gain vector $\mathbf{e}'_k, k{\in}\mathbf{K}$ as the ascending order of $\mathbf{e}_k$, while keeping the gain vector of other nodes as the descending order, the resultant final outcome ensures that each relay obtains larger rate than the sum rate of its descendant nodes.

- Another alternative is as follows: we obtain the gain vector of all nodes in the descending order, however we want to apply some tricks on matrix $\mathbf{\Gamma}$. First, we get a RB set $\mathbf{R}'$ with size $\sum_{k\in\mathbf{K}} n_k^*$ which contains best possible RBs for all relays. Then, we fix the rows of matrix $\mathbf{\Gamma}$ such that $U_{tj} = 0, t{\in}\mathbf{T}, j{\in}\mathbf{R}'$.

3.5.4 Performance Improvement using Power Control

Now, we are set with power and RB allocation $(\mathbf{x}^*, \mathbf{P}^*)$ for all nodes $t \in \mathbf{T} \cup \mathbf{K}$. For the tth node, the set of RBs is $I_{RB,t}$. The algorithms presented before subdivide power P_t among all RBs uniformly, $j \in I_{RB,t}$. Therefore, allocated power $p_{t,j}$ on RB j can be over-shot or under-shot. Hence, there is a scope of optimizing power for the settled tth node. We can formulate the problem as follows

$$\underset{\sum_j p_{t,j} = P_t}{\arg\max} \quad \chi_t(\mathbf{x}^*, P_t) \quad \forall t \in \mathbf{T} \cup \mathbf{K}. \tag{3.13}$$

The dual function for the problem in Equation 3.13 is $L_t(\mathbf{x}^*, \lambda_t)$. The solution of the dual problem is $\lambda_t^* = \arg\min_{\lambda_t > 0} L_t(\mathbf{x}^*, \lambda_t)$. A solution of the dual problem exists if and only if $\lambda_t = \frac{|I_{RB,t}|}{P_t + \sum_{j \in I_{RB,t}} \frac{1}{e_{t,j}}}$. This statement can be verified by the equation, $\sum_{j \in I_{RB,t}} \left(\frac{1}{\lambda_t} - \frac{1}{e_{t,j}} \right) - P_t = 0$. $|I_{RB,t}|$ is the length of set $I_{RB,t}$. Given λ_t^*, optimal power allocation for the problem in Equation 3.13 is $p_{t,j} = \frac{1}{\lambda_t^*} - \frac{1}{e_{t,j}}, \forall j \in I_{RB,t}$.

This type of problem and solution has already been given in [81] for the downlink tone allocation in OFDM-based systems. Their problem has an upper bound of SINR for each OFDM tone, whereas we do not have that restriction. They also have provided an iterative algorithm in order to obtain optimal λ_t. Since we do not have the SINR upper bound constraint, we have shown the modified algorithm in *Algorithm 5*. Before initiating the algorithm, we need to have one vector $\mathbf{a}_t$ which contains SINR $e_{t,j}$ for the RBs in set $I_{RB,t}$, i.e., $\mathbf{a}_t = \{e_{t,j}\}_{j \in I_{RB,t}}$. The complexity of this algorithm for the tth node is $|I_{RB,t}|$. Since there are $N_T + N_K$ number of nodes in the network, the overall complexity for this decision is $\sum_{t \in \mathbf{T} \cup \mathbf{K}} |I_{RB,t}|$.

In the first 3 subsections of this section, we have proposed the generic $\mathbf{\Lambda}$ update strategy (applicable to both SOA1 and SOA2), a low complexity suboptimal RB allocation algorithm with the help of original OFDMA tone allocation scheme, and relay oriented SOA2 algorithm which takes the assistance of the generic $\mathbf{\Lambda}$ update technique described in the first subsection. In the 4th subsection, we have given the optimal power allocation strategy of a node once it obtains RBs using the previously discussed methods.

3.6 Performance Evaluation

We have implemented relay enabled proposed algorithms in matlab and evaluate the performance of those comparing with the ones without relay [7] assuming the utility as both mere rate and the logarithmic of rate. In the following subsection, we describe how we configure the parameters of a network to perform our desired performance evaluation, and then in the next subsection, we show some results to justify our arguments.

3.6.1 Simulation Methodology

For setting up the network, we put the base station at the center, user nodes at different distance surrounding the base station. We also deploy some relay nodes in order to pass data from the edge nodes to the base station. We run the simulation over 200 TTIs. One TTI is equivalent to 1 ms and it consists of 50 RBs. Each RB is analogous to 12 subcarriers [6]. These total 50×12 subcarriers are spread over 10 MHZ bandwidth. The theoretical limit [82] of channel capacity is $\beta = \frac{-1.5}{\ln(5P_b)}$, where P_b is bit error rate (BER) and we configure it as 10^{-6}. For the user node, the maximum power is set as 220 mW and for a relay, it is moderately a large number depending on the number of terminal nodes it serves. The way how channel gain is computed is given in Equation 3.1. All results described in the following subsections

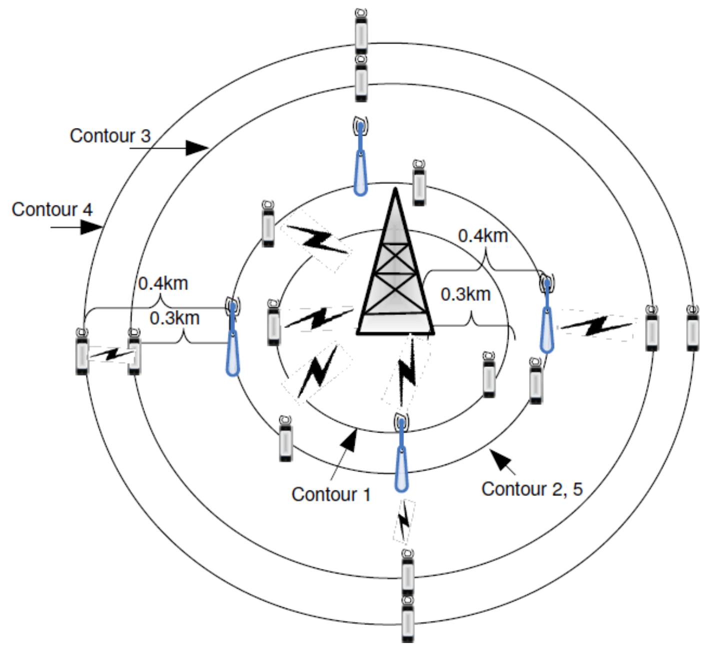


Figure 3.2: Simulation scenario (setup-1).

are the average of 20 simulation runs.

In the following subsection, we compare our algorithms based on utility function, $U = r$, with the original algorithms without relay [7] in terms of different performance metrics. Then, we compare the algorithms based on two utility functions, i.e., $U = r$ and $U = \ln(r)$. In all figures, for the utility function $U = r$, we denote *Algorithm 1*, *Algorithm 3* and *Algorithm 4* as $sopt(1)$, $sopt(2)$ and $sopt(3)$, respectively. On the other hand, for the utility function $U = \ln(r)$, we indicate them as $fsopt(1)$, $fsopt(2)$ and $fsopt(3)$.

The physical structure of the simulated cell looks like a circle. The base station is at the center of the circle. We have designed a few scenarios in order to evaluate the performance of our proposed algorithms by placing terminal nodes at different distance from the center of the circle. The Rayleigh fading effect is statistically similar for all nodes in the network. Therefore, the distance as well as the shadowing

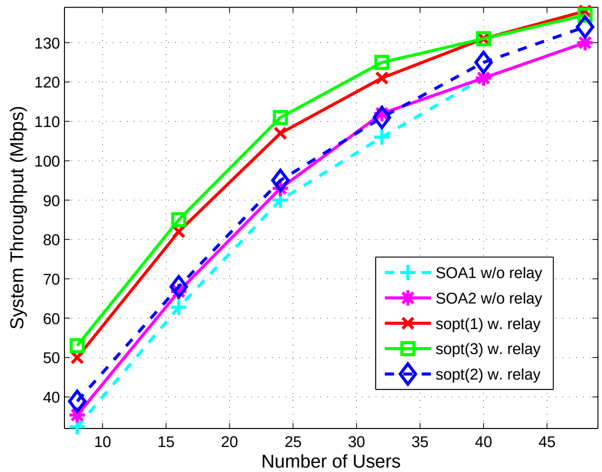
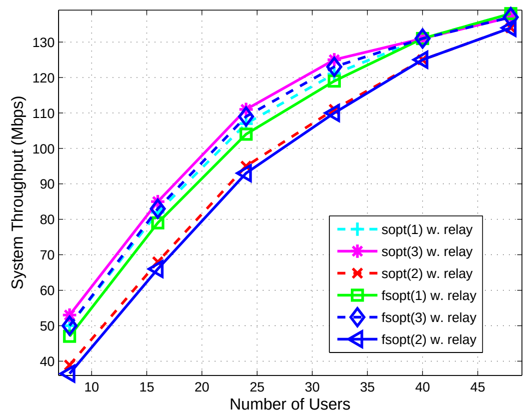


Figure 3.3: System throughput comparison among our relayed algorithms and those without relay (setup-1).

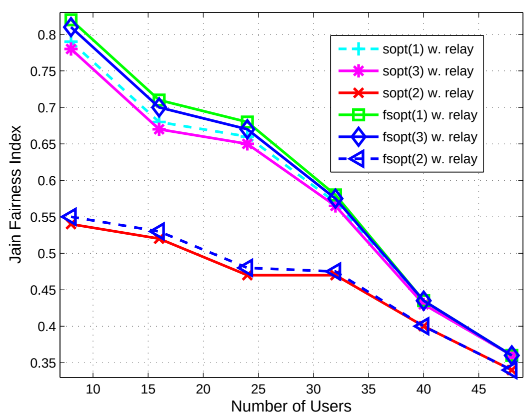
effect plays the pioneer role to distinguish the channel gain among them. Quite a lot scenarios are possible in terms of nodes placement in the network. It is not worth to show results of all scenarios, however we have justified the efficacy of our solutions by illustrating the results of some.

3.6.2 Simulation Results

First, we will show the results in terms of average system throughput, fairness, utilized number of RBs, number of active relays, number of terminals nodes served by the relays, etc. To measure the fairness quantitatively, we use Jain's fairness index algorithm [83]. This is a well-known metric used in network engineering to determine whether users or applications receive fair share of system resources. This metric is defined as follows



(a) System Throughput vs. Number of Users



(b) Jain Fairness Index vs. Number of Users

Figure 3.4: Comparison between throughput maximization and fair algorithms (setup-1).

$$\text{fairness}(r_1, r_2, \cdots, r_n) = \frac{\left(\sum_i r_i\right)^2}{\left(n.\sum_i r_i^2\right)},$$

where $r_1, r_2 \cdots$ represent individual nodes' throughput. Tentatively, we consider there are a few independent nodes and a few edge nodes in the network. We have placed some relays in order to help the edge nodes forwarding their data. For the purpose of clarity, we enumerate a few scenarios: (1) the channel condition of some or all relays can be statistically better/worse/equal than that of some or all independent nodes, (2) the channel condition of some or all edge nodes can be statistically better/worse/equal than that of some or all relays, etc. The channel gain of the edge nodes is computed considering its physical location with respect to its relay node. In all cases, our algorithms are adaptive, i.e., if we cannot accommodate relays due to the constraint of resource (power or RB), the scheduler does not count on those relays and assume their connected user terminals as the independent ones. Four scenarios we have presented, have 4 different contours at different distance around the base station on which we have placed our terminal nodes. The relay nodes are placed at the 5th contour. In the similar manner, as mentioned in the analytical section, we denote the number of terminal nodes as N_T. Among these nodes, $N_T/2$ is the number of independent nodes and the rest of them are edge nodes. Every 2 edge nodes, one from the 3rd contour and another one from the 4th contour are connected to 1 relay placed at the 5th contour. Therefore, the number of relays in the network is $N_K = N_T/4$[3].

[3]One of the assumptions of this work is that the routing information of the network is preconfigured. Hence, even if the nodes would be placed in a random manner, the route information of each node would need to be configured manually. To avoid this manual configuration, we would have followed some sort of design for the network setup.

For the first scenario, the contours of the independent nodes are $[0.3; 0.4]$ km away and the contour of the relays is 0.4 km away from the base station. Other two contours for the edge nodes are $[0.7; 0.8]$ km away from the base station, whereas from the designated relay, they are approximately $[0.3; 0.4]$ km away. Figure 3.2 depicts such system where the 5th contour is superimposed on the 2nd contour. Figure 3.3 illustrates gradual increment of system throughput with the increasing N_T. The more the number of users, the higher the throughput which is obvious in the figure. For using relays, our proposed solutions achieve higher throughput than when they are not used. As presented in [7], SOA2 has higher throughput than SOA1. Therefore, relay enhanced SOA2 has the highest performance except when the number of nodes in the system is 48, this is because $sopt(3)$ is more intolerant than $sopt(1)$ in achieving convergence and hence less number of relays remain active by the former one compared to the later one as we increase the number of nodes in the system. Moreover, with the increased number of terminal nodes, the number of relays is increasing as well. However, in each scheduling epoch, there is a limited number of RBs. In this scenario, the channel gain for half of the independent nodes is better comparing with the relay ones, the channel condition for half of the independent nodes is almost similar to the later ones. Moreover, maximum transmit power of the relays is larger than that of the regular terminal nodes. Therefore, when they have statistically similar channel gain, the relay nodes have higher priority in scheduling decision. At the first few data points, when the number of nodes is less compared to the number of available RBs, all of the relays stay active and all edge nodes connected to them are served properly because of the availability of RBs. Table 3.2 shows these findings of our scheduling algorithms. With the increased number of nodes, the scheduler cannot accommodate relays any more for the lack of resource and increased number of starved independent nodes. However, the channel condition and power of the relay nodes are better than the independent ones, we see, even for the number of nodes 48, $sopt(1)$ ensures some active relays and edge nodes served by them. If we would gradually increase N_T, we would see decremented number of relays as well as decremented number of edge nodes.

Table 3.1 shows the average number of used RBs in each scheduling epoch. This table proves that at low load, the scheduler leaves some unallocated RBs when relays are not used. This has motivated us to place some relays in order to serve edge nodes, which improves the performance of the edge nodes as well as the overall one. Moreover, precisely from the simulation, we have seen an interesting phenomenon when there are 3 remaining RBs unallocated and there are 3 edge nodes and 1 relay to serve them. Instead of subdividing 3 RBs among 3 edge nodes, giving 1 to the relay and other 2 to the edge nodes incur better performance comparing with the former one. At low load, even after deploying relays, all RBs are not used at each scheduling epoch. With the increased number of employed nodes, all RBs are utilized by either the terminal nodes or relays.

When the utility function is defined as mere rate, attained throughput is much larger comparing with the case when it is logarithmic of rate which has been depicted in Figure 3.4(a). However, if we look at Figure 3.4(b), fairness is much better with $U = \ln(r)$ than that with $U = r$. Thus, taking or not taking logarithmic function introduces the tradeoff between throughput maximization and fairness.

In the second scenario, the contours of the independent nodes are $[0.2; 0.3]$ km away from the base station. The positions of the relays and edge nodes are as same as in the previous scenario. The basic observations in Figure 3.5, such as increased throughput with increased number of terminal nodes, having better performance of

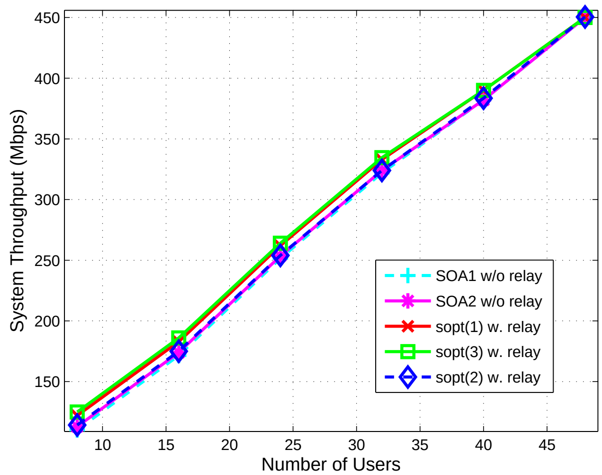


Figure 3.5: System throughput comparison among our relayed algorithms and those without relay (setup-2).

our algorithms with the deployed relays than that without relays, existence of tradeoff between throughput maximization and fairness due to two utility functions are as same as before. In this setup, the channel gain of all independent nodes is better than that of the relays. Therefore, with the increased number of nodes, the scheduler runs out of RBs and it subdivides available RBs among the nodes with better channel quality, i.e., independent nodes. As a result, we see, when the number of nodes is 48, no relays remain active and hence all edge nodes connected to the relays are treated as regular independent nodes. With *Algorithm 3*, we still see some active relays, this is because this algorithm does not try to reduce the duality gap. It just allocates resource among the independent nodes including relays based on the marginal utility. Since the maximum power of the relays is larger, some relays still have chance to have some RBs even when they have worse gain than the independent nodes from this scheduler's point of view. This scheduler gives privilege to the edge nodes if and only if their relays already have some allocation. On the other hand, our extended

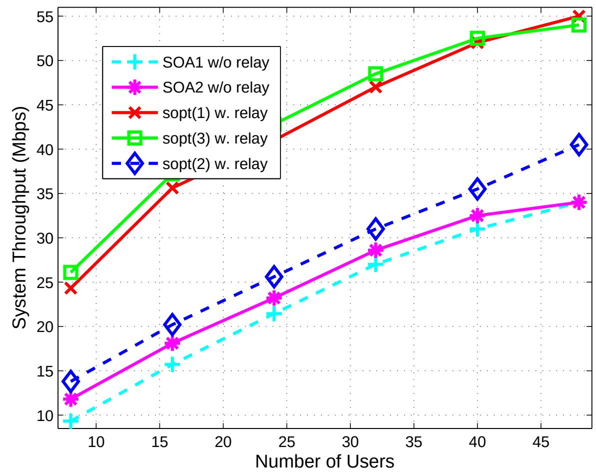


Figure 3.6: System throughput comparison among our relayed algorithms and those without relay (setup-3).

SOA1 and SOA2 go through a process of reducing duality gap for the constraint in Equation 3.4, and it becomes almost impossible when there is a constraint of resource with the increased number of terminal nodes. Because of these reasons, when the number of terminal nodes is large, *Algorithm 3* can accommodate a few relays as well as few edge nodes, and therefore it incurs slightly larger throughput comparing with others. This increased performance can be assumed as negligible. Whereas with our other heuristics, since the scheduler deactivates almost all relays, their performance reduces to the original ones. Table 3.2 is for the active relays and edge nodes for this scenario. Moreover, Table 3.1 reflects average number of used RBs resulting from this scenario.

The third scenario is the opposite of the second one. The contours of the independent nodes are placed at $[0.5; 0.6]$ km away from the base station, i.e., the channel gain of all relays is larger than theirs. Because of this reason, the relays always have more privilege in terms of RB allocation even when the number of nodes in the network

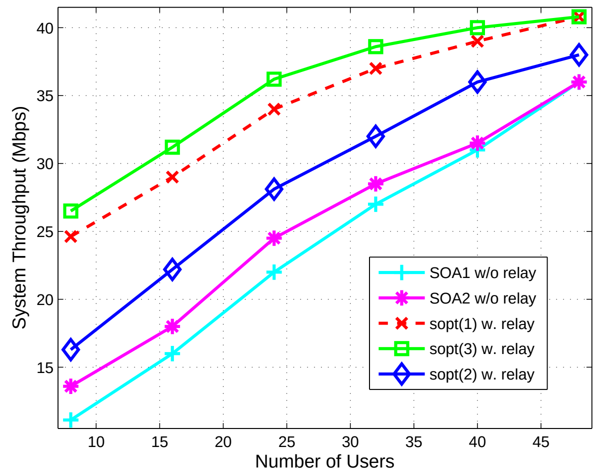


Figure 3.7: System throughput comparison among our relayed algorithms and those without relay (setup-4).

is large. Therefore, our main heuristics have more flexibility in reducing the duality gap for each relay and rearrange RBs among the relays and edge nodes. However, at increased load, the scheduler cannot fully ignore all independent nodes and it is hard to achieve convergence for all relays because of the resource constraint. Hence, to give more provision to the independent nodes, the scheduler deactivates some relays. Relay enhanced SOA2 is more intolerant in attaining convergence than *Algorithm 1*, which results in less number of active relays. Due to this reason, it has same or worse performance compared to the other one at some points which has been depicted in Figure 3.6. Whereas, *Algorithm 3* is ignorant and hence it has always worse performance than the main ones. Other basic observations for this scenario are as same as for the previous ones.

For the above all cases, all relays are feasible even at the beginning of iterations. We have prepared setup-4 in such a way that the attained rate of the relay nodes is smaller than the sum rate of their edge nodes. Therefore, according to the steps

[7 − 21] in *Algorithm 1*, *Algorithm 4*, to make it feasible, one edge node from all relays get disconnected and those nodes yield to be independent ones. The rest of the other settings of this setup are as same as the 3rd one. Hence, if we look at Table 3.2, we see, the number of active edge nodes is as same as the number of active relays even when the system has less number of nodes, and has sufficient resource to serve. Since the channel gain of the independent nodes is not that good compared to the relays and their edge nodes, the relays get more priority in consuming resource. Hence, we see, at low load, all relays remain active by serving their only one connected edge node. However, with the increased load, due to the lack of resource and starved independent nodes, more and more relays cannot be active. Since relay enhanced SOA2 is more intolerant, at higher load, it cannot serve as many relays as $sopt(1)$ can. Since $sopt(2)$ is ignorant of allocating resource for the independent nodes, relays and their connected edge nodes, at higher load, more relays are active whereas their connected nodes do not get any RB which is just the wastage of resource. The rest of the other observations are as same as for the previously discussed scenarios.

In order to see what happens to each single independent node because of the deployed relays, we have made a scenario with 8 terminal nodes and 1 relay. The 5th node is relay and it serves last 2 nodes. The distance of the nodes towards the base station is $[0.2; 0.3; 0.4; 0.45; 0.6; 0.7; 0.85; 0.9]$ km. Figures 3.8 and 3.9 illustrate average throughput and the number of allocated RBs for each single user, respectively. Most important observation from these 2 figures is, better independent nodes have lower throughput than the original ones because of taking away better RBs by the relay. Due to the employed relay, the best independent node has a lower number of RBs than the original ones. However, the employed relay improves the performance of the edge nodes drastically, whereas SOA1 and SOA2 cause starvation to them.

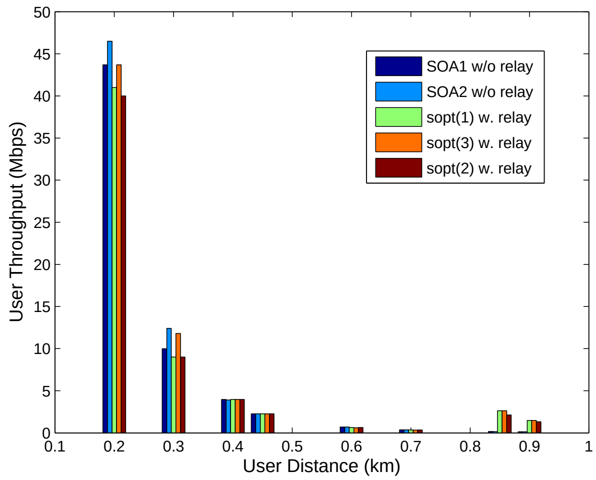


Figure 3.8: Individual user's throughput as a function of the distance from base station.

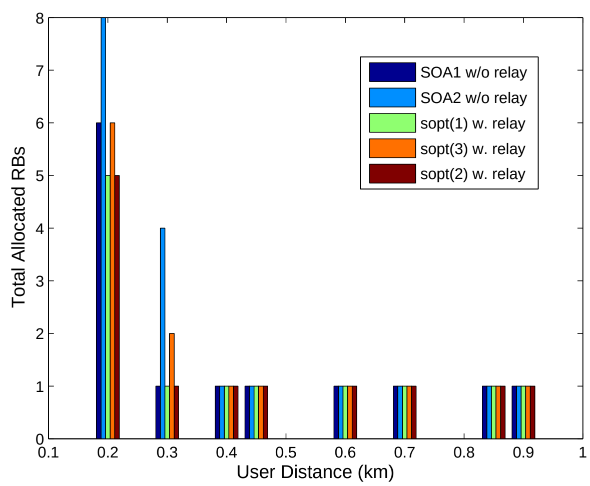


Figure 3.9: Individual user's allocated RBs as a function of the distance from base station.

SOA2 gives more provision to the nodes with better channel and so does the same relay enhanced SOA2 compared to $sopt(1)$. Because of the property in Figure 3.1, even with the deployed relay, all RBs in each scheduling epoch are not used.

3.7 Summary of the Chapter

In this work, we have presented the solution of the uplink scheduling for FD relay equipped LTE networks. First, we have shown optimization based formulation for our defined problem. From the dual formulation, we have noticed, we can use the sub-optimal algorithms proposed for OFDMA-based uplink scheduling problem without relay [7]. Based on this finding, we have proposed a family of suboptimal algorithms in order to perform resource allocation for our problem. Furthermore, for the fair resource allocation across the system, we have given a Nash bargaining solution which does not require much technical involvement on the top of the original formulation and its solution. Our proposed heuristics are adaptive, if due to the resource constraint, the relays do not contribute much to the real consumers in the system, they are recommended to be disabled. By designing some scenarios for actual systems, we have done extensive simulation on our proposed solutions and have proved that relays can enhance the performance of the system without consuming its obligatory resources.

Table 3.1: Total allocated RBs.

Alg/ # Users	SOA1	SOA2	sopt(1)	sopt(3)	sopt(2)
Setup-1					
8	16	18	17	18	12
16	24	30	31	32	20
24	32	38	40	42	31
32	40	44	46	48	39
40	44	46	49	50	45
48	50	50	50	50	50
Setup-2					
8	18	26	28	28	21
16	28	48	32	48	26
24	35	48	44	48	30
32	41	50	49	50	40
40	49	50	50	50	46
48	50	50	50	50	50
Setup-3					
8	10	10	12	14	11
16	18	20	20	32	19
24	28	30	31	42	29
32	36	40	40	48	38
40	45	50	50	50	46
48	50	50	50	50	50
Setup-4					
8	9	10	13	14	9
16	16	20	28	28	17
24	24	30	29	30	24
32	32	40	40	40	32
40	40	50	48	48	40
48	48	50	50	50	46

Table 3.2: Number of active relays/edge nodes.

Alg/ # Users	sopt(1)	sopt(3)	sopt(2)	fsopt(1)	fsopt(3)	fsopt(2)	Tot # Rlys/ Tot # Edges
Setup-1							
8	2/4	2/4	2/2	2/4	2/4	2/2	2/4
16	4/8	4/8	4/4	4/8	4/8	4/4	4/8
24	6/12	6/12	6/6	6/12	6/12	6/6	6/12
32	8/16	8/16	8/8	8/15	8/14	8/8	8/16
40	10/10	10/10	10/10	10/10	10/10	10/10	10/20
48	6/6	4/4	12/12	5/5	3/3	12/12	12/24
Setup-2							
8	2/4	2/4	2/2	2/4	2/4	2/2	2/4
16	4/8	4/8	4/4	4/8	4/8	4/4	4/8
24	6/12	6/12	6/6	6/12	6/12	6/6	6/12
32	8/14	8/12	9/8	6/6	8/12	8/10	8/16
40	10/10	10/10	11/10	9/9	8/8	10/10	10/20
48	0/0	0/0	8/7	0/0	0/0	8/7	12/24
Setup-3							
8	2/4	2/4	2/2	2/4	2/4	2/2	2/4
16	4/8	4/8	4/4	4/8	4/8	4/4	4/8
24	6/12	6/12	6/6	6/12	6/12	6/6	6/12
32	8/16	8/16	8/9	8/16	8/16	8/9	8/16
40	10/17	10/15	10/11	10/17	10/15	10/11	10/20
48	12/20	12/18	12/12	12/20	12/18	12/12	12/24
Setup-4							
8	2/2	2/2	2/2	2/2	2/2	2/2	2/4
16	4/4	4/4	4/4	4/4	4/4	4/4	4/8
24	6/6	6/6	6/6	6/6	6/6	6/6	6/12
32	8/8	8/8	8/8	8/8	8/8	8/8	8/16
40	9/9	9/9	10/10	9/9	9/9	10/10	10/20
48	8/8	8/8	12/10	8/8	8/8	12/10	12/24

Algorithm 1 Iterative algorithm to determine $(\mathbf{x}^*, \mathbf{P}^*)$.

for $k \in \mathbf{K}$ **do**
    $\mathbf{F}(t) := \{k\}, \forall t \in \mathbf{T}_k$.
end for
$\mathbf{V} := \{\infty\}, \boldsymbol{\theta} := \{0\}, u := 1, \boldsymbol{\Lambda}(u) := \{0\}$.
repeat
    Inner Loop for obtaining $(\mathbf{x}^*, \mathbf{P}^*)$.
    for $\forall k \in \mathbf{K}$ **do**
        if $N_{T_k} > 1$ **then**
            if $\sum_{t \in \mathbf{T}_k} g_t^b(j) > g_k^b(j)$ **then**
                $t^* := \arg\max_{t \in \mathbf{T}_k} g_t^b(j)$.
                $\mathbf{F}(t^*) := 0$.
            else
                $\theta(k) := 1$.
            end if
        else
            if $\sum_{t \in \mathbf{T}_k} g_t^b(j) > g_k^b(j)$ **then**
                $\mathbf{F}(k) := -2$.
            end if
            $\theta(k) := 1$.
        end if
    end for
    if $\forall k \in \mathbf{K}(\theta(k) = 1)$ **then**
        if $u = 1$ **then**
            $\kappa(k) := \frac{\epsilon}{|\sum_{t \in \mathbf{T}_k} g_t^b(j) - g_k^b(j)|}, \forall k \in \mathbf{K}$.
        end if
        for $\forall k \in \mathbf{K}$ **do**
            $\mathbf{F}(t) := 0, \theta(k) := 0, \exists t \in \mathbf{T}_k g_t^b(j) = 0$.
            if $\sum_{t \in \mathbf{T}_k} g_t^b(j) = 0$ OR $(g_k^b(j) = 0$ AND $\exists t \in \mathbf{T}(\mathbf{F}(t) = 0, g_t^b(j) = 0))$ **then**
                $\mathbf{F}(k) := -3, \theta(k) := 0$.
            end if
            Mark the kth relay if its convergence is achieved.
        end for
        if $\sum_{t \in \mathbf{T}_k} g_t^b(j) \approx g_k^b(j), \forall k \in \mathbf{K}$ **then**
            Convergence has been achieved.
        else if $\exists k \in \mathbf{K} \theta(k) = 0$ **then**
            Initialize u, $\mathbf{V}$ and $\boldsymbol{\Lambda}$.
        end if
        Follow *Algorithm 2*.
    end if
until Λ^* is found

Algorithm 2 Intermediate fraction of *Algorithm 1*.

for $\forall k \in \mathbf{K}$ **do**
    if the kth relay is not converged **then**
        $\Lambda_k(u) := \Lambda_k(u-1) + \kappa(k) * |\sum_{t \in \mathbf{T}_k} g_t^b(j) - g_k^b(j)|$.
        $rf := \mathbf{F}(k) = -2?1 : -1$.
        $\mathbf{V}(k) := g_k^b(j) - (0 - rf)\Lambda_k(u) g_k^b(j)$.
        $\mathbf{V}(t) := g_t^b(j) + (0 - rf)\Lambda_t(u) g_t^b(j), \forall t \in \mathbf{T}_k$.
    end if
end for
Remove disabled relays from sets $\mathbf{K}$ and $\mathbf{F}$.

Algorithm 3 Suboptimal RB allocation algorithm.

$j := 0$, $\Omega_i(j) = \emptyset$ for each terminal or relay i.
for $k \in \mathbf{K}$ **do**
    $\mathbf{F}(t) = k, \forall t \in \mathbf{T}_k$.
end for
while $j < N_{RB}$ **do**
    $j := j + 1$.
    Update RB index $l_i(j)$ for each node i.
    Calculate metrics $g_i^b(j)$ and $g_i^a(j)$ for each node i.
    Set $\xi := \{i | \mathbf{F}(i) <= 0\}$.
    for $i \in \{i | \mathbf{F}(i) > 0\}$ **do**
        if $\left(\sum_{t \in \mathbf{T}_k, t \neq i} g_t^b(j) + g_i^a(j)\right) <= g_{\mathbf{F}(i)}^b(j)$ **then**
            $\xi := \xi \cup i$.
        end if
    end for
    Find $i^* := \arg\max_{i \in \xi}(g_i^a(j) - g_i^b(j))$.
    if $(g_{i^*}^a(j) - g_{i^*}^b(j)) <= 0$ **then**
        Break the loop.
    end if
    Assign the j'th RB to node i^*

$$\Omega_i(j) = \begin{cases} \Omega_i(j-1) \cup l_i(j) & \text{if } i = i^* \\ \Omega_i(j-1) & \text{Otherwise} \end{cases}.$$

end while

Algorithm 4 SOA2 extended RB allocation algorithm.

1: Perform steps $[1-4]$ of *Algorithm 1*.
2: Solve problem (10) and obtain resultant $U_t^*(\lambda^*), r_t^*(\lambda^*), t \in \mathbf{T} \cup \mathbf{K}$.
3: Map $r_t^*(\lambda^*)$ to $g_t^b(j), t \in \mathbf{T} \cup \mathbf{K}$.
4: Perform steps $[7-39]$ of *Algorithm 1* which requires to repeat steps $[2-3]$ until $\mathbf{\Lambda}^*$ is obtained.
5: Apply *Hungarian Algorithm* as described in order to achieve $\mathbf{C}^*$, i.e., $(\mathbf{x}^*, \mathbf{P}^*)$.

Algorithm 5 Iterative algorithm for optimal λ_t.

[a] = sort(a, 'descend').
$j := 0, G_n = 0, G_d = 0$.
$j := j + 1$.
while $j <= |I_{RB,t}|$ **do**
    $G_n := G_n + 1,$.
    $G_d := G_d + \frac{1}{a_t(j)}$.
    $\lambda_t(j) = \frac{G_n}{P_t + G_d}$.
    if $\lambda_t(j) >= a_t(j+1)$ **then**
        $\lambda_t^* = \lambda_t(j)$.
        Break the loop.
    end if
end while

Chapter 4

Joint Source and Relay Power Allocation for AF Relayed Network (Centralized Solution)

4.1 Introduction

Emerging relay technique in 3G/4G networks, such as LTE and LTE-Advanced has brought numerous problems to optimize its deployment. In recent years, cooperative relay networks have received tremendous attention. Two most popular cooperation protocols are considered - DF, AF [84]. DF relays retransmit just the replica of source's transmitted signal. Whereas AF means, the relay nodes amplify source's signal first, and then retransmit towards the destination. We adopt the AF protocol in our work because of its simple signal processing mechanism. Our problem presented in this chapter focus on a network having one relay node, a set of sources, and single destination. Given fixed amount of power to be distributed among the sources and relay is an optimization problem, and exhibits the trade off between fair resource allocation and overall network performance. If resource is limited in the network, serving all sources may lead to the degradation of network capacity. In this situation, best possible few sources should be selected for the transmission while ignoring others.

To put our work into context, we outline the chronological order of evolved research on the power optimization of AF relayed networks. At the beginning of such exploration, people focused on a simple network with one source and one relay [39]. And, then, relay power allocation for a single-source multi-relay network has extensively been studied varying different optimization criterion, i.e., outage probability minimization or sum relay power minimization under the sources' SNR or outage probability constraints [40], [41], [42]. Performance optimization of multi-source single-relay network has also been given attention from different angles, such as interference cancellation schemes are proposed in [43] and [44]. In [85] and [45], network decoding is applied to combat interference among the users. In [86], the resource including both subcarrier and relay power allocation problem is studied to maximize the sum rate. At the same time, joint multi-source multi-relay power optimization has appeared in a few papers [46], [47] under different optimization criterion, i.e., sum rate maximization, sum power minimization, etc. Although in [46], each source is essentially served by only one relay, [47] made performance improvement by assisting single source with the help of multiple relays. The drawback of their approach is, they have totally ignored SNR due to the direct link transmission in their optimization formulation.

Due to the inefficient bandwidth utilization of relays while transmitting on the orthogonal channels, at one point, many people gave emphasis on the best relay selection schemes. Multi-relay selection and their power allocation recently appeared in some works [87], [88], [89] because of their ability to attain better performance. Joint relay and opportunistic source selection for a bidirectional network has been considered in [90] to optimize the outage probability and BER.

In this chapter, we have formulated the problem considering individual source, relay, and total system power constraints. Total power constraint is imposed to reduce the interference in the neighboring network. Our formulated problem is almost similar to [46], [47], although [46] has not taken individual source and total power constraints, and in [47], each source is served by multiple relays[4]. The main drawback of their work is, they have totally skipped SNR due to the direct link in their formulation. Putting the direct link SNR in the original formulation, we have solved the problem using GP. Because of the direct link SNR, the original problem is not amenable to GP. Hence, we have used single condensation method in order to solve this problem using GP. Since the original problem has many variables, and single condensation method may need many iterations to attain convergence, we have transformed our problem to the problem with two variables (one source and one relay) which is much computationally cheap. Once we obtain the source transmit power, we subdivide it among all original sources using greedy and fair algorithms. For dividing the relay power among them, we apply water filling method given that the sources are assigned with transmit power using the proposed heuristics.

The rest of the chapter is organized as follows. We explain the system model and its detailed mechanism in Section 4.2. In Section 4.3, we provide the centralized solution of our defined problem. We evaluate the performance of this solution in Section 4.4, and finally we conclude the chapter in Section 4.5.

4.2 System Model and Problem Formulation

In this section, we consider an LTE system, in which there exists a base station, one relay node and N source nodes which need help from the relay to get their packets

[4]For the sake of simplicity without losing the generality, we consider, the sources are served by a single relay node.

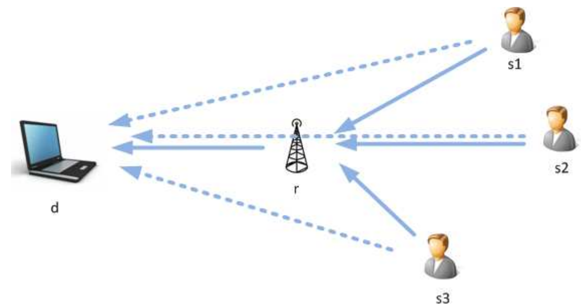


Figure 4.1: System model.

transmitted to the base station. The edge nodes essentially act as servers which have some special applications and the destination node needs to get the content of that application on time. For example, the application could be some video which needs to be displayed on the destination. The relay node amplifies the received signal from the source nodes and then forwards towards the destination as depicted in Figure 4.1. Although the system has one destination, the proposed solution can easily be adapted to a multi-destination network. The transmit channel of each source or the relay is a single or aggregate RB(s). We assume a block-fading channel (or quasi-static) model: the channels remain invariant over a time slot whose duration is less than the coherence time of the channels. Denote the channel gain between source s_i and destination d is $G_{s_i,d}$; the channel gain between source s_i and relay r is $G_{s_i,r}$; and the channel gain between relay r and destination d is $G_{r,d}$.

The entire transmission operation using the AF relay consists of two phases (i.e., time slots). At each phase, the sources or relay use orthogonal RBs for multiple transmissions. At the first phase, source s_i broadcasts its information to both destination d and relay node r. The received signals $y_{s_i,d}$ and $y_{s_i,r}$ at destination d and relay r can be expressed as

$$y_{s_i,d} = \sqrt{E_{s_i} G_{s_i,d}} x_{s_i} + \eta_{s_i,d} \text{ and } y_{s_i,r} = \sqrt{E_{s_i} G_{s_i,r}} x_{s_i} + \eta_{s_i,r}, \tag{4.1}$$

where E_{s_i} represents the transmit power at node s_i, x_{s_i} is the broadcast information symbol with unit energy from source s_i to nodes d and r. $\eta_{s_i,d}$ and $\eta_{s_i,r}$ are the additive noises received at destination d and relay r, respectively. In the second step, the relay amplifies its received signal and forwards it to destination d. Denote the power the relay uses to help source s_i is E_{r_i}. The signal received at destination d for source s_i can be shown

$$y_{r_i,d} = \frac{\sqrt{E_{r_i} G_{r,d}} \left(\sqrt{E_{s_i} G_{s_i,r}} x_{s_i} + \eta_{s_i,r} \right)}{\sqrt{E_{s_i} G_{s_i,r} + \sigma^2}} + \eta_{r_i,d}. \tag{4.2}$$

$\eta_{r_i,d}$ is the received noise from relay r to destination d (for source s_i). Without loss of generality, we assume that the noise power is the same additive white gaussian noise for all links, denoted by σ^2. After maximum ratio combining of both the direct and relay paths, the effective received SNR for source s_i's transmission can be given by

$$\Gamma_{s_i,r,d} = \frac{E_{s_i} G_{s_i,d}}{\sigma^2} + \frac{E_{s_i} G_{s_i,r} E_{r_i} G_{r,d}}{\sigma^2 \left(E_{s_i} G_{s_i,r} + E_{r_i} G_{r,d} + \sigma^2 \right)}. \tag{4.3}$$

If the set consisting of the source nodes is $L_s = \{s_1, s_2, \cdots, s_N\}$, the total capacity achieved by the system can be given by

$$R_{s,r,d} = \gamma_L W \sum_{s_i \in L_s} \log_2 \left(1 + \Gamma_{s_i,r,d} \right). \tag{4.4}$$

Because the transmissions are orthogonal, $\gamma_L = 1/(2N)$ and W is the aggregate bandwidth in the system. Since W and γ_L are constants, we skip these terms in the subsequent discussion.

Our goal is to allocate power among the sources and relay so that the system capacity is maximized. As in traditional network resource optimization problems, there are constraints on the sources and relay power. Moreover, in order to mitigate the interference imposed on another network due to the transmission operations in this network, there is a total power constraint, meaning total power allocated to the sources and relay node cannot exceed E^{max}. For the sake of simplicity, we have converted the maximization problem into the minimization one by introducing minus sign in front of the objective function, i.e, $R_{s,r,d}$.

$$\min \prod_{s_i \in L_s} \frac{\sigma^2(\sigma^2 + E_{s_i} G_{s_i,r} + E_{r_i} G_{r,d})}{(\sigma^2 + E_{s_i} G_{s_i,d})(\sigma^2 + E_{s_i} G_{s_i,r} + E_{r_i} G_{r,d}) + E_{s_i} G_{s_i,r} E_{r_i} G_{r,d}} \tag{4.5}$$
$$\text{where} \quad E_{s_i} \leq E_s^{max}, \quad s_i \in L_s,$$
$$\sum_{s_i \in L_s} E_{r_i} \leq E_r^{max},$$
$$\sum_{s_i \in L_s} E_{s_i} + \sum_{s_i \in L_s} E_{r_i} \leq E^{max},$$
$$\{E_{s_i}\}_{s_i \in L_s} \geq 0, \ \{E_{r_i}\}_{s_i \in L_s} \geq 0.$$

The aforementioned optimization problem is valid if and only if $\sum_{s_i \in L_s} E_{s_i} + E_r^{max} > E^{max}$ and $\sum_{s_i \in L_s} E_{s_i}^{max} > E^{max}$.

4.3 Centralized Solution

The problem in Equation 4.5 is not convex due to the non-convexity property of the objective function. This statement can be proved very easily by the help of special type of convex optimization formulation, i.e., GP [91, 92][5]. A GP is a type of mathematical optimization problem characterized by the objective and constraint

[5]The objective function is the ratio of two posynomials, which is non-convex.

functions that has a special form. It focuses on monomial and posynomial functions. A monomial is a function, $h : \mathbb{R}^n \rightarrow \mathbb{R}$, where the domain contains all real vectors with non-negative components, $h(x) = c{x_1}^{a_1}{x_2}^{a_2} \cdots {x_n}^{a_n}$. A posynomial is a sum of monomials, $f(x) = \sum_k c_k {x_1}^{a_{1k}}{x_2}^{a_{2k}} \cdots {x_n}^{a_{nk}}$. GP is an optimization problem with the form

$$\text{minimize} \;\; f_0(x) \;\; \text{subject to} \;\; f_i(x) \leq 1, \; h_j(x) = 1,$$

where f_0 and f_i are posynomials and h_j are monomials. This problem in the above form is not convex. However, with a change of variables: $y_i = \log x_i$ and $b_{ik} = \log c_{ik}$, we can transform it into convex form given the assumption that the logarithm of a sum of exponentials is a convex function.

As mentioned, the objective function is the ratio of two posynomials which cannot be solved by GP. There are ways to transform such type of problem to GP form, i.e., single condensation method, double condensation method [92]. We have used single condensation method which requires to approximate the denominator of the objective function by some monomial term. We denote the denominator by $F\left(\{E_{s_i}\}_{s_i \in L_s}, \{E_{r_i}\}_{s_i \in L_s}\right)$ and the resultant monomial is given by

$$\prod_{s_i \in L_s} (\sigma^2 + E_{s_i}G_{s_i,d})(\sigma^2 + E_{s_i}G_{s_i,r} + E_{r_i}G_{r,d}) + E_{s_i}G_{s_i,r}E_{r_i}G_{r,d} \approx \lambda \prod_{s_i \in L_s} E_{s_i}^{a_i} E_{r_i}^{b_i},$$

where $a_i = \frac{E_{s_i}}{F\left(\{E_{s_i}\}_{s_i \in L_s}, \{E_{r_i}\}_{s_i \in L_s}\right)} \frac{\partial F\left(\{E_{s_i}\}_{s_i \in L_s}, \{E_{r_i}\}_{s_i \in L_s}\right)}{\partial E_{s_i}}$,
$b_i = \frac{E_{r_i}}{F\left(\{E_{s_i}\}_{s_i \in L_s}, \{E_{r_i}\}_{s_i \in L_s}\right)} \frac{\partial F\left(\{E_{s_i}\}_{s_i \in L_s}, \{E_{r_i}\}_{s_i \in L_s}\right)}{\partial E_{r_i}}$,
and $\lambda = \frac{F\left(\{E_{s_i}\}_{s_i \in L_s}, \{E_{r_i}\}_{s_i \in L_s}\right)}{\prod_{s_i \in L_s} E_{s_i}^{a_i} E_{r_i}^{b_i}}$.

The derivations are below

$$
\begin{aligned}
&\frac{\partial F\left(\{E_{s_i}\}_{s_i\in L_s},\{E_{r_i}\}_{s_i\in L_s}\right)}{\partial E_{s_i}} = \\
&[G_{s_i,d}(\sigma^2+E_{s_i}G_{s_i,r}+E_{r_i}G_{r,d})+(\sigma^2+E_{s_i}G_{s_i,d})G_{s_i,r}+G_{s_i,r}E_{r_i}G_{r,d}] \\
&\prod_{s_j\in L_s, s_j\neq s_i}\left[(\sigma^2+E_{s_j}G_{s_j,d})(\sigma^2+E_{s_j}G_{s_j,r}+E_{r_j}G_{r,d})+E_{s_j}G_{s_j,r}E_{r_j}G_{r,d}\right],
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial F\left(\{E_{s_i}\}_{s_i\in L_s},\{E_{r_i}\}_{s_i\in L_s}\right)}{\partial E_{r_i}} = \\
&[G_{r,d}(\sigma^2+E_{s_i}G_{s_i,d})+E_{s_i}G_{s_i,r}G_{r,d}] \\
&\prod_{s_j\in L_s, s_j\neq s_i}\left[(\sigma^2+E_{s_j}G_{s_j,d})(\sigma^2+E_{s_j}G_{s_j,r}+E_{r_j}G_{r,d})+E_{s_j}G_{s_j,r}E_{r_j}G_{r,d}\right].
\end{aligned}
$$

Finally, the overall procedure for the joint source and relay power allocation is given as follows.

1. Set the initial value of power $\mathbf{E}^{(0)} := \left[E_{s_1}^{(0)}, \cdots, E_{s_N}^{(0)}, E_{r_1}^{(0)}, \cdots, E_{r_N}^{(0)}\right]$, $n := 1$.
2. Determine $\left[a_1^{(n)}, \cdots, a_N^{(n)}\right]$, $\left[b_1^{(n)}, \cdots, b_N^{(n)}\right]$ and $\lambda^{(n)}$.
3. Solve the optimization problem with the help of GP.
4. Denote the optimal power allocation in the nth round as $\mathbf{E}^{(n)}$.
5. If $||\mathbf{E}^{(n)} - \mathbf{E}^{(n-1)}|| \leq \epsilon$, where ϵ is a pre-defined threshold, the enumerations stop; otherwise, $n := n+1$ and reiterate from step 2 to 5.

The above procedure updates $2N$ principal variables in every iteration. Each iteration needs to update $2N+1$ number of intermediate variables to assist updating the principal variables. In order to simplify this procedure, we can consider, the system has only one source node denoted by s^*. Node s^* is the representative of all sources. The gain between node s^* and node r is the weighted average of the gains between the sources and relay. In the similar manner, the gain between node s^* and node d is determined. After this transformation, the objective function is still the ratio of two posynomials. In order to cast it to GP, we can approximate the denominator (denoted by $H(E_{s^*}, E_r)$) of it by the monomial

$$\begin{aligned}&(\sigma^2 + E_{s^*}G_{s^*,d})(\sigma^2 + E_{s^*}G_{s^*,r} + E_rG_{r,d}) + E_{s^*}G_{s^*,r}E_rG_{r,d}\\ &\approx \mu E_{s^*}^c E_r^d,\end{aligned}$$

where $c = \frac{E_{s^*}}{H(E_{s^*},E_r)}\frac{\partial H(E_{s^*},E_r)}{\partial E_{s^*}}$, $d = \frac{E_r}{H(E_{s^*},E_r)}\frac{\partial H(E_{s^*},E_r)}{\partial E_r}$, and $\mu = \frac{H(E_{s^*},E_r)}{E_{s^*}^c E_r^d}$. Furthermore,

$$\begin{aligned}\frac{\partial H}{\partial E_{s^*}} &= G_{s^*,d}(\sigma^2 + E_{s^*}G_{s^*,r} + E_rG_{r,d}) + (\sigma^2 + E_{s^*}G_{s^*,d})G_{s^*,r} + G_{s^*,r}E_rG_{r,d},\\ \frac{\partial H}{\partial E_r} &= (\sigma^2 + E_{s^*}G_{s^*,d})G_{r,d} + E_{s^*}G_{s^*,r}G_{r,d}.\end{aligned}$$

The iterative procedure used to obtain optimal E_{s^*} and E_r follows the same procedure mentioned above. However, it requires to update two principal variables and three auxiliary variables in each iteration to achieve convergence. The following two subsections are for distributing power $E_{s^*}^*$ among all original sources, and the third subsection is for disseminating relay power E_r^*.

4.3.1 Suboptimal Source Power Allocation (Greedy Solution)

From the optimal solution, we have observed that the sources with better channel condition obtain more power compared to others. Since each source has individual power constraint and this power is moderately lower than the total allowable power for all sources, we can propose a greedy power allocation for the source nodes given the total allowable for them is $E^*_{s^*}$. If the direct link SNR of a source is better than its relayed link one, it is likely, that source obtains zero relay power. Therefore, it is rational to distribute $E^*_{s^*}$ among the sources taking direct link SNR into account. We sort $G_{s_i,d}, s_i \in L_s$ in decreasing order and allocate the maximum individual power to each sorted source until there is no left over power.

The above approach for distributing power among the sources is greedy. This is similar to the MaxCIR technique [93] which assigns subcarrier among the users in OFDM-based networks according to their channel condition. The drawback of this approach is, the sources with worse channel may starve and may never get chance to transmit as they are assigned zero power. This reminds us one important issue, which is called fairness. In order to tackle fairness, we have proposed an algorithm which considers both the instantaneous channel condition and fairness.

4.3.2 Suboptimal Source Power Allocation (Fair Solution)

At scheduling time instant t, we denote the gains between the sources and relay as a vector $\boldsymbol{\alpha}_t = \{\alpha_t(1), \alpha_t(2), \cdots, \alpha_t(N)\}$; the gains between the sources and destination as a vector $\boldsymbol{\gamma}_t = \{\gamma_t(1), \gamma_t(2), \cdots, \gamma_t(N)\}$; the gain between the relay and destination as β_t. Moreover, the average rate of the sources is denoted by a vector $\bar{\boldsymbol{\zeta}}_t = \{\bar{\zeta}_t(1), \bar{\zeta}_t(2), \cdots, \bar{\zeta}_t(N)\}$. We initialize all elements of this vector as 0 at time

$t = 0$. A fair algorithm is presented in *Algorithm 6*[6]. The steps in the algorithm are followed in the channel coherent time of each time instant t.

Algorithm 6 Fair algorithm for subdividing transmit power among the sources.

```
Get α_t, γ_t and β_t.
Sort γ_t in the descending order and get sorted index set I := {I_1, I_2, ··· , I_N}.
Set χ := E*_{s*}.
for ∀i∈I do
    if ζ̄_{t-1}(i) = 0 then
        η(i) := min(E_s^{max}, χ).
        χ := χ − η(i).
    end if
end for
if χ > 0 then
    Let the last unallocated index j := i.
    for ∀j∈I do
        Set M(j) := γ_t(j) / ζ̄_{t-1}(j).
    end for
    Sort index vector I according to the descending order of vector M.
    for ∀j∈I do
        Set η(j) := min( χ*M(j) / Σ_{j=i}^{N} M(j), E_s^{max} ).
        χ := χ − η(j).
    end for
end if
```

4.3.3 Suboptimal Relay Power Allocation

In order to subdivide relay power E_r^* among all sources, we have adopted water filling approach and the resultant formulated problem is given by

$$\underset{\sum_{s_i \in L_s} E_{r_i} = E_r^*}{\arg\max} \sum_{s_i \in L_s} \log_2 \left(1 + \frac{E_{s_i} G_{s_i,r} E_{r_i} G_{r,d}}{\sigma^2 (E_{s_i} G_{s_i,r} + E_{r_i} G_{r,d} + \sigma^2)} \right). \tag{4.6}$$

[6]Here, the fairness metric is the ratio of the user's instantaneous rate in time instant t and its past average throughput.

By invoking the Lagrange multiplier μ for the total relay power constraint of the problem in Equation 4.6, we obtain the Lagrangian $\sum_{s_i \in L_s} \log_2 \left(1 + \frac{E_{s_i} G_{s_i,r} E_{r_i} G_{r,d}}{\sigma^2 (E_{s_i} G_{s_i,r} + E_{r_i} G_{r,d} + \sigma^2)}\right) + \mu(E_r^* - \sum_{s_i \in L_s} E_{r_i})$. Following the K.K.T condition, we take the differentiation of the Lagrangian with respect to E_{r_i}, and we obtain

$$\begin{gathered}\mu\sigma^2(\sigma^2 + E_{s_i} G_{s_i,r} + E_{r_i} G_{r,d})^2 + \mu E_{s_i} G_{s_i,r} E_{r_i}^2 G_{r,d}^2 + \\ \mu(\sigma^2 + E_{s_i} G_{s_i,r}) E_{s_i} G_{s_i,r} E_{r_i} G_{r,d} - (\sigma^2 + E_{s_i} G_{s_i,r}) E_{s_i} G_{s_i,r} G_{r,d} = 0.\end{gathered} \tag{4.7}$$

After simplifying, the resultant E_{r_i} is $-\frac{E_{s_i} G_{s_i,r} + 2\sigma^2}{2G_{r,d}} + \frac{1}{2G_{r,d}} \sqrt{E_{s_i}^2 G_{s_i,r}^2 + \frac{4E_{s_i} G_{s_i,r} G_{r,d}}{\mu}}$. Substituting $E_{r_i}, s_i \in L_s$ back into the equation $\sum_{s_i \in L_s} E_{r_i} - E_r^* = 0$, we obtain the upper bound of μ, which is $\frac{\sum_{s_i \in L_s} \sqrt{4E_{s_i} G_{s_i,r} G_{r,d}}}{2G_{r,d} E_r^* + \sum_{s_i \in L_s} (2\sigma^2 + E_{s_i} G_{s_i,r})}$. As the lower bound of μ is 0, we apply a bisection search between these two bounds in order to obtain optimal μ. Replacing the optimal μ in E_{r_i}, finally we obtain optimal E_{r_i}, i.e., $E_{r_i}^*, s_i \in L_s$.

4.4 Performance Evaluation

In this subsection, we will evaluate the performance of our proposed solutions. Section 4.4.1 is for the methodology we have adopted to evaluate the performance and the following one presents the results while comparing with the approach proposed in [46] and [47].

4.4.1 Simulation Methodology

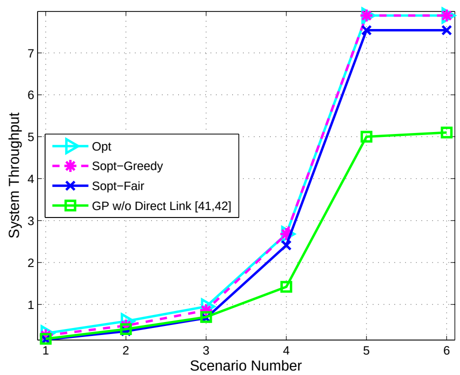
We presume an LTE network with 5 source nodes, one relay and one destination node, i.e, base station. The maximum power of individual source is $E_s^{max} = 30$ mW, and that of the relay node is $E_r^{max} = 50$ mW, the total available power in the system is $E^{max} = 120$ mW. Noise variance σ^2 has been set as 1. The channel between two nodes suffers from the shadowing and Rayleigh fading effects. We take the same channel model and the similar values of its parameters as mentioned in Chapter 3. Moreover, we assume, each channel has a unit capacity. One of the major assumptions of the works [46], [47] is, the channel condition between the source and destination is always worse than that between the source and relay. However, that not necessarily happens in practice, and for the counter scenario, their model fails to provide optimal solution. In order to fix the model up, we have considered the SNR due to the direct link in our formulation, and the resultant solution is optimal which is able to give better performance even when the direct link's channel is better than that of the relayed link. In order to evaluate the performance of our solution, we have selected 6 different scenarios, each of which has 5 distinct source nodes. In the evaluation part, we have denoted each scenario by Scenario Number. The positions of all nodes are in the following coordinates:

- Destination: (0,12).
- Relay: (0,6).
- Sources: X-coordinates are fixed at $\{-1, -2, -3, -4, -5\}$ for all scenarios. If Scenario Number is denoted by n_s, their Y-coordinates is $2n_s$, if $n_s \in \{1, 2, 3, 4, 5\}$; and $2n_s$, if $n_s = 6$.

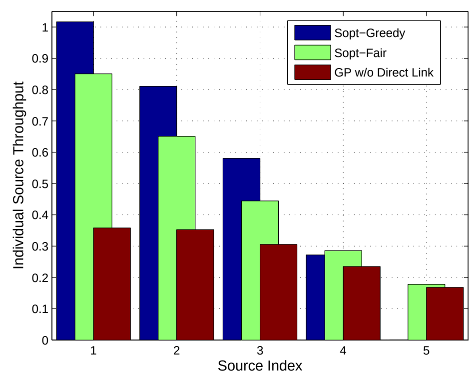
All the results we have presented here are the average of 100 simulation runs.

4.4.2 Simulation Results

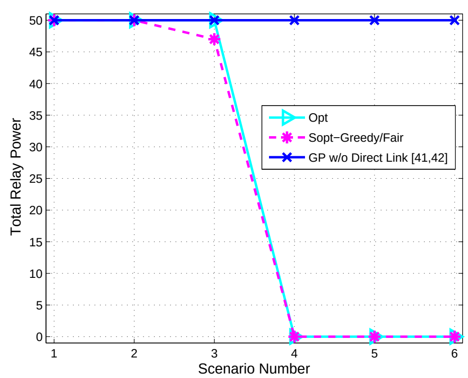
Figure 4.2(a) presents the total average system throughput with respect to 6 different scenarios of the source nodes. Notice that the positions of the relay and destination



(a) System throughput.



(b) Individual source throughput.



(c) Total relay power.

Figure 4.2: Comparison between our centralized solutions and others.

Table 4.1: Source power comparison between optimal and suboptimal greedy solutions.

Scenario Number	s_1		s_2		s_3		s_4		s_5	
	Opt	Greedy	Opt	Greedy	Opt	Greedy	Opt	Greedy	Opt	Greedy
1	28.29	30	26.46	30	15.24	10	0	0	0	0
2	29.12	30	28.21	30	12.66	10	0	0	0	0
3	30.0	30	30	30	10	12	0	0	0	0
4	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	0	0
5	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	0	0

Table 4.2: Source relay power comparison between optimal and suboptimal solutions.

Scenario Number	s_1		s_2		s_3		s_4		s_5	
	Opt	Sopt	Opt	Sopt	Opt	Sopt	Opt	Sopt	Opt	Sopt
1	36.19	35.01	13.78	14.98	0	0	0	0	0	0
2	36.96	35.91	13.031	14.08	0	0	0	0	0	0
3	32.24	31.12	17.75	16.84	0	0	0	0	0	0

are fixed, we are varying the positions of 5 sources towards the destination. As the sources move to the destination, the resultant channel gain becomes better for them, hence gradually their throughput get improved. For the 5th and 6th scenarios, the absolute distance between the sources and destination is close, however the sources are on the different sides of the destination. Since their absolute distances are close, the resultant throughput are close for these two scenarios. Now if we intuitively compare all approaches, for scenarios 1 and 2, the direct link's channel condition is worse than the relayed link, the resultant outcome proposed by [46], [47] does not deviate much from our optimal solution. For scenario 3, the channel quality of the direct link is close to the relayed link and from this scenario, the procedure without considering the direct link SNR starts to differ from our optimal approach. And, for scenarios 4, 5 and 6, channel of the relayed link is worse than that of the direct link. For these scenarios, Figure 4.2(c) shows that allocated power for the relay is 0, and the total allowable power is distributed among the sources considering their channel gain towards the destination. However, the technique without considering the direct link SNR always assigns full power to the relay no matter the relayed link is worse or better than the direct link. Since our suboptimal approach for allocating power to the source and relay is based on somewhat weighted averaging of all gains, for scenario 3, relayed power by this approach is little less than the optimal one. For rest of the other scenarios, the suboptimal approach confers to the optimal one. Table 4.1 and Table 4.2 compare the detailed breakdown of power allocation between the optimal and suboptimal schemes. We have noticed that once we obtain the total allowable power for the sources, we can distribute this power among the sources by 2 techniques, i.e., greedy and fair algorithms. From Figure 4.2(a), the greedy one has close performance comparing with the optimal one. Greedy algorithm gives privilege to the sources with better channel condition, and makes starvation for others. For being fair to the sources with worse channel condition, the resultant system throughput by the fair algorithm deteriorates compared to the other one. Even for scenarios 1, 2, and 3, achieved throughput by this algorithm is worse compared to the GP solution without considering the direct link SNR, however for the other scenarios, it outperforms.

Figure 4.2(a) has sources at different distance from the relay as well as from the destination. In order to have detailed performance comparison of these four approaches, for scenario 4, we have shown each individual source's throughput contribution towards the overall performance of the system in Figure 4.2(b). In the X-axis, we put source node index and in the Y-axis, the corresponding node's throughput contribution has been projected. As discussed in the previous paragraph, the relay power is 0 for this scenario. Therefore, the technique without considering the direct link SNR has worse performance except some fluctuation comparing with the fair one. The greedy algorithm first assigns full allowable power to the source with the best channel, and this process goes on for all sources with better channel quality until the total allowable power runs out. Because of this nature of power subdivision, source 5 obtains 0 power since its channel condition is the worst compared to the rest others. The fair algorithm assigns some power to the sources with worse channel over the time, and hence those sources contribute some throughput towards the overall performance. Because of giving some privilege to this type of sources, the sources with the best channel quality obtain less amount of power compared to that obtained by the greedy one. Therefore, with the fair algorithm, the source node with the best channel has worse performance compared to the greedy one.

4.5 Summary of the Chapter

In this chapter, we have studied both transmit and relay power allocation problem in a multi-source single-AF-relay network. Since the existing works [46], [47] of this problem have not considered the direct link SNR in their formulation, the resultant solution deviates from the optimal one if the direct link SNR is better than the relayed link one. Having noticed this, we put the missing term in the original formulation, and have solved this using GP. Because of incorporating the direct link SNR in the problem formulation, the problem is not directly amenable to GP. Hence, we have used single condensation method in order to cast it to GP. Since this solution is computationally expensive with the growing number of sources, we also have given two suboptimal solutions. The suboptimal solutions are designed considering the greedy and fair nature of the source nodes. Extensive simulation results verify that the proposed suboptimal solutions achieve close performance of the optimal one.

Chapter 5

Joint Source and Relay Power Allocation for AF Relayed Network (Game Theoretical Solution)

5.1 Introduction

In the aforementioned solution of our defined problem in Chapter 4, the nodes in the network are assumed to be altruistic, and willing to cooperate in optimizing the overall network performance. In many practical solutions, however, nodes are selfish and aim to optimize their own benefits or utilities, which may result in conflict of sharing resource among them. Game theory is a flexible and natural tool to model, and analyze these behaviors of the nodes. This tool has extensively been applied for several problems in cooperative relay networks. In [94], uplink of a network with multiple users can form a coalition in order to maximize their transmission rates or utilities. In [95], [96], a bargaining game has been designed in order to show the interaction between two users, where one user acts as a relay for the other one with the purpose of attaining bandwidth [95] and power allocation [96]. In [97], the users define the payment rates for the relays, and they share the payments among themselves who are willing to help the users. [98] shows cooperation for two nodes in two different mesh networks. An analytical framework is proposed to determine when the cooperation is beneficial, and the expected performance gain is estimated.

Game theoretical power allocation for a multi-user multi-relay network has also extensively been studied. For example, non-cooperative game theory has been applied to show the competition among the relays in order to assign optimal power.The authors in [99] have presented a solution for relay power allocation of such network, where two kinds of users are considered: variable, constant rate users. They have used dual decomposition method in order to solve their formulated problem, and based on that, they have given a distributed solution. For a single-user multi-relay network, a two level Stackelberg game has been proposed in [100] for selecting the best relays and determine their power. For a multi-user single-relay network, source power allocation problem has also been modeled using a two level Stackelberg game [101]. The relay sets the price for the users, and they play a non-cooperative game in order to maximize their individual utility. One more recent work for a multi-source single-relay network is [102] for the purpose of relay power allocation following some fair resource allocation rule. The relay acts as a leader and sets the price for power, whereas the users work as customers or followers for the purpose of maximizing their utilities. In this work, since the relay needs to know the complete information of the network, the same authors proposed a fully distributed relay power allocation scheme for the users [103].

Having observed the non-convexity nature of our problem, and to model the selfish nature of the sources, in this chapter, we have proposed a game theoretical solution with two steps[7]. Two Stackelberg games operating in two consecutive steps have been designed in order to solve this problem. For connecting these two games,

[7]Since the joint sources and relay power allocation of this network is a non-convex problem, it is impossible to capture this problem with a best-response-strategy-based single game.

we have introduced one centralized entity, which is aware of the complete channel state information (CSI) in the network. In the first step, this entity plays as a buyer level game for buying power from the source nodes. The sources are non-cooperative among themselves in terms of selling their power to the centralized entity. On the other hand, the centralized entity is willing to maximize its own utility by settling the optimal amount of power when the prices are announced by the sources. In the second step, on behalf of the relay node, the centralized entity plays as a seller level game for selling/disseminating relay power to the sources. In turn, the sources themselves compete for the relay power given the price set by the centralized entity. As there is a total power constraint in the system, at the beginning of the game, the centralized entity applies some intelligent measure in order to determine how much power is dedicated for source transmission and how much is for relay operation[8]. Although there are some works on source power control [101], and on relay power allocation [102] for a multi-source single-relay network, there is no complete solution for joint sources and relay power allocation for such network in the literature, which we believe to fill up. Moreover, the work in [101] considers that the source nodes transmit simultaneously in the same frequency/time domain which results in interference among them, and not a practical notion of a relayed system. Through extensive simulation while showing the results from different stand points, we have justified that the game theoretical solution achieves comparable performance with the centralized one.

The rest of the chapter is organized as follows. System model with its mechanism is briefly summarized in Section 5.2. Game theoretical solution of defined problem

[8]Except the centralized entity, no other nodes need to know the CSI of other nodes. At the beginning of these games, the centralized entity uses the complete CSI to decide the amount of power for the transmit and relay operations in aggregate level. However, two games decide the amount of power for the individual source. In each game, the interaction between the sources and the centralized entity is completely distributed.

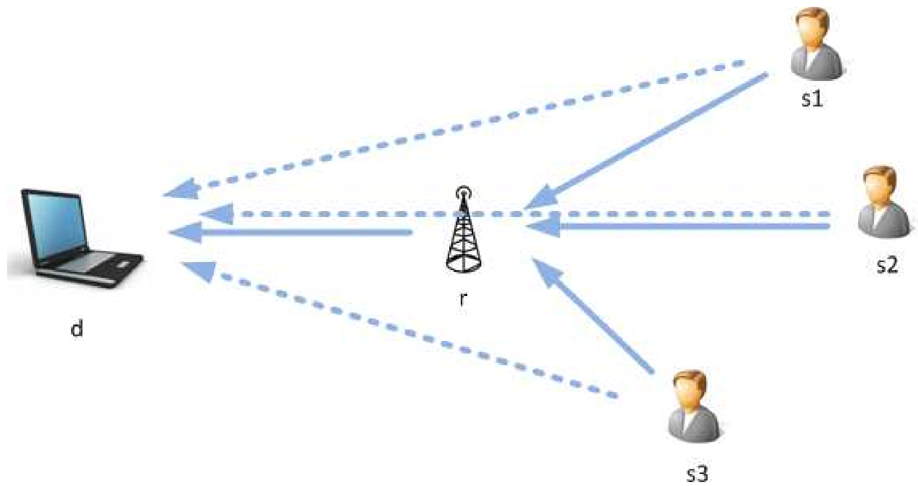


Figure 5.1: System model.

is given in Section 5.3. We evaluate the performance of this solution in Sections 5.4 and compare the performance with the centralized optimal solution given in Chapter 4. Finally, Section 5.5 concludes the chapter.

5.2 System Model and Problem Formulation

For the system shown in Figure 5.1, detailed description of the problem is given in Chapter 4. Our objective is to solve the following optimization problem

$$\max \sum_{s_i \in L_s} \log_2 \left(1 + \frac{E_{s_i} G_{s_i,d}}{\sigma^2} + \frac{E_{s_i} G_{s_i,r} E_{r_i} G_{r,d}}{\sigma^2 \left(E_{s_i} G_{s_i,r} + E_{r_i} G_{r,d} + \sigma^2\right)}\right) \tag{5.1}$$

$$\begin{gathered} \text{where} \quad E_{s_i} \le E_s^{max}, \quad s_i \in L_s, \\ \sum_{s_i \in L_s} E_{r_i} \le E_r^{max}, \\ \sum_{s_i \in L_s} E_{s_i} + \sum_{s_i \in L_s} E_{r_i} \le E^{max}, \\ \{E_{s_i}\}_{s_i \in L_s} \ge 0, \ \{E_{r_i}\}_{s_i \in L_s} \ge 0, \end{gathered}$$

where the channel gain between source s_i and destination d is $G_{s_i,d}$; the channel gain between source s_i and relay r is $G_{s_i,r}$; and the channel gain between relay r and destination d is $G_{r,d}$. $E_{s_i}^{max}$ is the maximum power constraint of source s_i and E_r^{max} is that of relay r. For mitigating the interference in the neighboring network, the maximum allowable power E^{max} is imposed in this network.

5.3 Game Theoretical Solution

Since joint sources and their relay power allocation is a non-convex problem, by employing a single cooperative or non-cooperative game, it is not possible to assign transmit and relay power of a source jointly.[9] Hence, we plan to consider sources and their relay power allocation as two distinct problems. First, putting some assumption on the relay power distribution, the sources will decide their optimal power independently. In the next step, in order to improve the performance of the network further, optimal relay power distribution for the sources is decided. However, in order to connect these two problems, we need an entity in the network whom we call "Network". Primarily, Network is aware of the complete CSI of the sources. It is also aware of the individual source, relay, and total system power constraints.

In the centralized solution of this problem, all nodes in the network work selflessly to maximize the network capacity. However, in the real world, selfish nodes may not have a common goal or belong to a common authority. Therefore, a reimbursement mechanism is required such that the sources can earn some benefits while contributing power towards maximizing the capacity of the network. Since there is a restriction

[9] If we would formulate the problem with only one best-response-strategy-based game, we would not be able to design any cost function which is convex with respect to both transmit and relay power of each source. However, this is the fundamental requirement of formulating a best-response-strategy-based game: the cost function has to be convex with respect to its variable (transmit power, relay power or price).

on the total power taken from the sources, the authority node "Network" is likely to choose the most beneficiary sources whose contribution are better compared to others. Following the characteristics of the sources and Network, a Stackelberg game is appropriate to model this problem while considering their joint benefits. Network plays the buyer level game since it aims to achieve the best performance from power given by the sources while giving the least possible reimbursement to them. On the other hand, each source plays the seller level game, in which it aims to earn the payment that not only covers its power cost but also gains as much extra profit as possible.

In order to improve the performance of the network further, Network wants the optimal distribution of relay power among the sources. However, due to the selfish nature, the sources may want to have as much power as possible to maximize their own SNR. Furthermore, relay power is limited, if one source takes the whole power, it results in starvation for others. Given the total relay power budget, the objective of the sources is non-cooperative. To discourage the sources having as much power as they want, pricing is an effective mechanism. If price is imposed on the relay power, they will ask for the amount of power which maximizes their individual utility in order to keep the balance between power and price. In order to model such behaviors of the sources for their relay power, we have introduced second a Stackelberg game. In this game, Network is the entity that will decide the price of unit relay power. Given the unit power price, the sources demand some power which maximizes their own utility or benefit.

In order to ensure the correct and unique convergence of the first game, Network needs to know how much total power is dedicated for the transmit operation of the sources. The rest of the power subtracted from the total network power is for the relay operation. Following the formula given below, the total relay power of the sources is determined. For the first game, if the relay power is not 0, Network assumes that available relay power is subdivided equally among the sources which need the help of relay.

$$s_i \in \begin{cases} L_s^1 & \text{if } G_{s_i,d} < \frac{G_{s_i,r}G_{r,d}}{(G_{s_i,r}+G_{r,d})} \\ L_s^2 & \text{Otherwise} \end{cases}, \tag{5.2}$$

$$E'_r = \begin{cases} 0 & \text{if } L_s^1 \text{ is empty} \\ E_r^{max} & \text{Otherwise} \end{cases}. \tag{5.3}$$

If L_s^1 is non-empty, $E'_{r_i} = E'_r/|L_s^1|, s_i \in L_s^1$. $|L_s^1|$ is the length of set L_s^1.

5.3.1 Game Theoretical Source Power Allocation

This game is to identify the number of sources served by Network, and how much power to be disseminated among them given that total allowable power is E'_s and the maximum allowable power for source s_i is $E_{s_i}^{max}$. We formulate the game below.

1. Network/Buyer: Network can be configured as a buyer, and it aims to attain most benefits at the least possible payment. So, the utility function of Network can be defined as

$$U_n^s = aR'_{s,r,d} - P, \tag{5.4}$$

where $R'_{s,r,d}$ is the aggregate SNR for all sources $\sum_{s_i \in L_s} \left[\frac{E_{s_i}G_{s_i,d}}{\sigma^2} + \frac{E_{s_i}G_{s_i,r}E'_{r_i}G_{r,d}}{\sigma^2\left(\sigma^2 + E_{s_i}G_{s_i,r} + E'_{r_i}G_{r,d}\right)} \right]$[10].

[10]Since log is a concave function, for the sake of simplicity, we have ignored log in the formulation.

a denotes the gain per unit of SNR achievement, P refers to the total payment paid by Network to the source nodes, $P = \sum_{s_i \in L_s} p_i E_{s_i}$, and p_i is the price per unit of power selling from source s_i to Network, and E_{s_i} represents the amount of power Network would like to buy from source s_i when the prices are advertised by them. Suppose, the sources preferred by Network consists of a set, represented by $L_s = \{s_1, s_2, \cdots, s_N\}$, then the optimization problem for Network can be given by

$$\max_{\{E_{s_i}\}_{s_i \in L_s}} U_n^s = aR'_{s,r,d} - P, \ s.t. \ E_{s_i} \geq 0, s_i \in L_s. \tag{5.5}$$

2. Sellers/Sources: Each source s_i can be assumed as a seller, and wants to earn the payment which not only covers its cost due to the contribution towards the overall system performance but also achieve as much extra profit as possible. Introducing one parameter $c_i, s_i \in L_s$, the cost of unit power which is a reflection of the sources' perception about whether they can actually get profit, the utility function of source s_i can be defined as

$$U_{s_i} = p_i E_{s_i} - c_i E_{s_i} = (p_i - c_i) E_{s_i}, \tag{5.6}$$

where E_{s_i} is source s_i's outcome by optimizing U_n^s illustrated in Equation 5.5. It is important to note that optimal p_i depends not only on source s_i's channel condition towards the destination and relay, but also on the prices of its partner sources. So, in each round of the game, if one source asks a higher price than what Network anticipates, after jointly comparing prices of all sources and their potential contribution to the overall performance, Network will buy less power from that source or even overlooks that source. On the other hand, if the price is too low, the profit obtained by Equation 5.6 will be uselessly low. Therefore, there is a tradeoff in setting the price. The optimization problem for source s_i or the game of seller s_i is

$$\max_{\{p_i\}>0} U_{s_i} \;=\; (p_i - c_i)E_{s_i}. \tag{5.7}$$

The fundamental purpose of above game is to decide the optimal price p_i for source s_i to maximize its profit U_{s_i}; the actual number of sources who will finally be selected by Network, and the corresponding power consumption $E_{s_i}, s_i \in L_s$ to maximize U_n^s. The following subsection shows the outcome of the game in detail.

Analysis of Source Power Allocation Game

We first examine the proposed game in detail, and obtain the closed form outcome of the game. Based on the outcome, we establish that obtained solution is the unique equilibrium called the Stackelberg Equilibrium (SE) of the game. Then, from the properties of the game, in the following subsection, we outline a distributed price update function, and the interaction mechanism of the entities.

(i) Network/Buyer Level Analysis: Before taking decision about the amount of power it will buy from the sources, it is crucial to know the prices asked by them. Taking the partial derivative of U_n^s with respect to E_{s_i}, from Equation 5.5, we obtain

$$\frac{\partial U_n^s}{\partial E_{s_i}} \;=\; a\frac{\partial R'_{s,r,d}}{\partial E_{s_i}} - p_i, \quad s_i \in L_s.$$

For U_n^s being strictly concave with respect to $\{E_{s_i}\}_{s_i \in L_s}$, condition $\frac{\partial U_n^s}{\partial E_{s_i}} > 0, s_i \in L_s$ should be satisfied. This means, p_i of the source should satisfy $p_i < a\frac{\partial R'_{s,r,d}}{\partial E_{s_i}}$.

At the beginning, source s_i has knowledge of its cost c_i, which is the bare expense required for its contribution towards the overall system performance, however is unaware of the prices of other sources. In order to get its utility U_{s_i} non-negative, at the first iteration, source s_i sets its price $p_i = c_i$. If under this lowest initial price, Network is reluctant to buy power from that source, s_i will not exist in the game anymore.

Consequently, before initiating the game, Network applies some intelligence in order to sort out the sources which it will play the game with. At first, Network tentatively set $E_{s_i} = 0, s_i \in L_s$, if for some source, say s_i, it holds that $c_i \geq \left(a\frac{\partial R'_{s,r,d}}{\partial E_{s_i}}\right)|_{E_{s_i}=0}$, s_i will be disregarded by Network.

For the remaining sources in set L_s, by the first order optimality condition, the following equation must be satisfied at the optimal point.

$$\frac{\partial U_n^s}{\partial E_{s_i}} = 0, \quad s_i \in L_s. \tag{5.8}$$

Solving Equation 5.8, we can get its solution $\{E^*_{s_i}\}_{s_i \in L_s}$ shown in *Lemma 5.1.*

Lemma 5.1: The optimal power consumption by source $s_i \in L_s$ depends on the contents of the sets L_s^1 and L_s^2 which were determined by Network at the beginning of the game.

Case 1: L_s^1 is non-empty, and L_s^2 is empty.

$$E^*_{s_i}(p_i) = \begin{cases} 0 & p_i \geq p_i^{ub} \\ \sqrt{\frac{aB_{s_i}E'_{r_i}G_{s_i,r}G_{r,d}}{G^2_{s_i,r}\sigma^2(p_i - \frac{aG_{s_i,d}}{\sigma^2})}} - \frac{B_{s_i}}{G_{s_i,r}} & p_i^{lb} < p_i < p_i^{ub} \\ E_s^{max} & \frac{aG_{s_i,d}}{\sigma^2} < p_i \leq p_i^{lb} \\ \text{Undefined} & p_i \leq \frac{aG_{s_i,d}}{\sigma^2} \end{cases},$$

where $B_{s_i} = E'_{r_i} G_{r,d} + \sigma^2$, $p_i^{lb} = \frac{aB_{s_i} E'_{r_i} G_{s_i,r} G_{r,d}}{\sigma^2 \left(E_s^{max} G_{s_i,r} + B_{s_i}\right)^2} + \frac{aG_{s_i,d}}{\sigma^2}$ and $p_i^{ub} = \frac{aE'_{r_i} G_{s_i,r} G_{r,d}}{B_{s_i} \sigma^2} + \frac{aG_{s_i,d}}{\sigma^2}$.

Case 2: L_s^1 is empty, and L_s^2 is non-empty.

Optimal power $E^*_{s_i}, s_i \in L_s^2$ of this case is obtained by solving the following optimization problem. In order to contribute non-negation utility towards the overall performance, the price lower and upper bounds of the sources $s_i \in L_s^2$ are $p_i^{lb} = 0$ and $p_i^{ub} = \frac{aG_{s_i,d}}{\sigma^2}$, respectively.

$$\underset{\{E_{s_i}\}_{s_i \in L_s^2}}{\arg\max} \sum_{s_i \in L_s^2} \frac{aE_{s_i} G_{s_i,d}}{\sigma^2} - \sum_{s_i \in L_s^2} p_i E_{s_i}. \tag{5.9}$$

From the solution of this optimization problem, we observe that full power E_s^{max} is assigned to the sources following the descending order of the metric $\frac{aG_{s_i,d}}{\sigma^2} - p_i, s_i \in L_s^2$ until the total allowable power E'_s runs out.

Case 3: Both L_s^1 and L_s^2 are non-empty.

This case is the hybrid scenario of above Case 1 and Case 2. Similar to the solution method of Case 2, we define metric $A_{s_i}, s_i \in L_s^1 \bigcup L_s^2$ for the sources in this case.[11] The physical meaning of A_{s_i} is the profit of source s_i for the given price p_i.

$$A_{s_i} = \begin{cases} aG_{s_i,d} - p_i & s_i \in L_s^1 \\ aG_{s_i,d} + \frac{aG_{s_i,r} E'_{r_i} G_{r,d}}{(G_{s_i,r} + E'_{r_i} G_{r,d})} - p_i & s_i \in L_s^2 \end{cases}. \tag{5.10}$$

We sort $A_{s_i}, s_i \in L_s^1 \bigcup L_s^2$ in the descending order. Then, following the order, we assign power among the sources in the following manner until the total allowable power E'_s runs out.

[11] When the direct link SNR is better than the relayed link one, relay service is not used for that case and this makes different types of functions for the profit of different sources.

$$E^*_{s_i}(p_i) = \begin{cases} E_s^{max} & s_i \in L_s^1 \\ \min\left(E_s^{max}, E_{s_i^*}(p_i)\right) & s_i \in L_s^2 \end{cases}. \tag{5.11}$$

The last allocated source may not get the full power following the formula defined above. It obtains the left over power from E'_s after assigning among the sources which have relatively better value for metric A_{s_i}.

After Network announces optimal power $(E^*_{s_i})^+$ to the sources $s_i \in L_s$, they will gradually increase the prices $p_i, s_i \in L_s$ to get possibly more benefit round by round. This will lead Network to buy decreasing amount of E_{s_i}. In order to earn the maximal utility instead of being disregarded by Network, source s_i also needs to ask proper price.

(ii) Source/Seller Level Analysis: Replacing the output from *Lemma 5.1* into Equation 5.7, we have

$$\max_{\{E_{s_i}\}>0} U_{s_i} = (p_i - c_i) E^*_{s_i}(p_i). \tag{5.12}$$

Notice that this game among the sources is non-cooperative, and there exists a tradeoff between the price p_i and utility U_{s_i} of source s_i. There is an optimal price to set for all sources in order to avoid being disregarded by Network, and maximize its own utility. The optimal price depends upon the source's channel condition as well as its own price. Network only chooses the most beneficiary sources for meeting up its own interest. Following the first order optimality condition, it results in

$$\frac{\partial U_{s_i}}{\partial p_i} = E^*_{s_i}(p_i) + (p_i - c_i)\frac{\partial E^*_{s_i}(p_i)}{\partial p_i} = 0, \quad s_i \in L_s. \tag{5.13}$$

Solving Equation 5.13, we obtain the optimal price $p_i^* = p_i^*(G_{s_i,r}, G_{r,d}, G_{s_i,d}, \sigma^2), \forall s_i \in L_s$.

The solutions in *Lemma 5.1* and p_i^* are an equilibrium of each round in this game. The properties and convergence procedure of the equilibrium are illustrated in the following subsection.

Properties of Source Power Allocation Game

In this subsection, we prove the existence of an SE in this game, and prove the optimality of the SE by the following properties.

Property 5.1: The utility function of Network U_n^s is concave with respect to $\{E_{s_i}\}_{s_i \in L_s}$, where $E_{s_i} \geq 0$, $\forall s_i \in L_s$, when the prices of the source nodes are constant.

Proof: Taking the second order derivatives of U_n^s, we get

$$\frac{\partial^2 U_n^s}{\partial E_{s_i}^2} = -2a \frac{(E'_{r_i} G_{r,d} + \sigma^2) E'_{r_i} G_{s_i,r}^2 G_{r,d}}{(E_{s_i} G_{s_i,r} + E'_{r_i} G_{r,d} + \sigma^2)^3}, \forall s_i \in L_s^1,$$

and

$$\frac{\partial^2 U_n^s}{\partial E_{s_i} \partial E_{s_j}} = 0, \quad s_i, s_j \in L_s^1.$$

From the above 2 equations, it is pretty much straightforward that $\frac{\partial^2 U_n^s}{\partial E_{s_i}^2} \frac{\partial^2 U_n^s}{\partial E_{s_j}^2} - (\frac{\partial^2 U_n^s}{\partial E_{s_i} \partial E_{s_j}})^2 > 0, \forall s_i \neq s_j$. Furthermore, U_n^s is continuous with respect to E_{s_i}. So, when $E_{s_i} \geq 0$, U_n^s is strictly concave in $\{E_{s_i}\}_{s_i \in L_s^1}$ and jointly concave as well. For Case 2, when L_s^1 is empty, U_n^s is non-differentiable with respect to $\{E_{s_i}\}_{s_i \in L_s^2}$, and the second derivative of U_n^s with respect to any $s_i \in L_s^2$ is 0. This concludes the concavity property of U_n^s with respect to $\{E_{s_i}\}_{s_i \in L_s^2}$ for Case 2. Case 3 is the hybrid scenario of both Cases 1 and 2. Since U_n^s is concave for both Cases 1 and 2, it is straightforward to say that U_n^s is concave with respect to E_{s_i} for Case 3.

Property 5.2: The optimal power consumption E_{s_i} has a decreasing trend with p_i when the prices of other sources are some fixed quantity.

Proof: Taking the first order derivative, we have

$$\frac{\partial E^*_{s_i}(p_i)}{\partial p_i} = -\frac{1}{2\sigma G_{s_i,r}}\sqrt{\frac{aB_{s_i}E'_{r_i}G_{s_i,r}G_{r,d}}{(p_i - \frac{aG_{s_i,d}}{\sigma^2})^3}} < 0, \tag{5.14}$$

which implies that $E^*_{s_i}$ is decreasing with p_i. For Case 2, if the prices of other sources are constant, increment of p_i drives Network to buy non-increasing amount of power from source s_i. In other way, we can say that if one source increases its price while other sources keep their prices constant, Network will buy less power from that source.

Property 5.3: The utility U_{s_i} of source s_i is concave in terms of its price p_i it asks for, given that its power consumption is the optimized amount demanded from Network as calculated in *Lemma 5.1* and also the prices of other sources are some fixed quantity.

Proof: $E^*_{s_i}(p_i)$ is a continuous function of p_i. Since U_{s_i} is a function of $E^*_{s_i}(p_i)$ and p_i, U_{s_i} is continuous in p_i. Taking the derivatives, we obtain

$$\frac{\partial U_{s_i}}{\partial p_i} = E^*_{s_i}(p_i) + (p_i - c_i)\frac{\partial E^*_{s_i}(p_i)}{\partial p_i}, \tag{5.15}$$

$$\frac{\partial^2 U_{s_i}}{\partial p_i} = 2\frac{\partial E^*_{s_i}(p_i)}{\partial p_i} + (p_i - c_i)\frac{\partial^2 E^*_{s_i}(p_i)}{\partial {p_i}^2}, \tag{5.16}$$

where

$$\frac{\partial^2 E^*_{s_i}(p_i)}{\partial {p_i}^2} = \frac{3\sqrt{B_{s_i}}}{4\sigma G_{s_i,r}}\sqrt{\frac{aE'_{r_i}G_{s_i,r}G_{r,d}}{(p_i - \frac{aG_{s_i,d}}{\sigma^2})^5}}.$$

We know that $p_i > c_i$. Furthermore, we have observed in Case 1 that if $p_i < \frac{aG_{s_i,d}}{\sigma^2}$, Network buys undefined power. Therefore, $c_i > \frac{aG_{s_i,d}}{\sigma^2}$ and the following statement is true.

$$\sqrt{\frac{1}{(p_i - \frac{aG_{s_i,d}}{\sigma^2})^3}} > \sqrt{\frac{1}{\frac{(p_i - \frac{aG_{s_i,d}}{\sigma^2})^5}{(p_i - c_i)^2}}}.$$

Because of the above statement and U_{s_i} is continuous with respect to p_i, from Equation 5.16, we can conclude that $\frac{\partial^2 U_{s_i}}{\partial p_i} < 0$, which justifies the concavity property of U_{s_i} with respect to p_i for Case 1. For other cases, it is straightforward to conclude the concavity property of U_{s_i} with respect to p_i.

Remark 5.1: Source Selection procedure by Network described in Section 5.3.1(i) is sufficient.

Proof: This is because, any source is rejected by Network at the beginning by this rule, however mistakenly taken by Network to play the game with, eventually that source will be rejected. The proof of this statement is as follows. Suppose that the source rejection criterion is applied for some source node s_i, i.e., $\frac{\partial U_n^s}{\partial E_{s_i}} < 0$, when $E_{s_i} = 0$ and $p_i = c_i$. And, Network does not exclude source s_i and in the following price update process, all source nodes gradually increase their prices to get more utilities. To prove that the new resulting $E^*_{s_i}(c_i + \delta) < 0$, it suffices to prove that $\sum_{s_j \in L_s} \frac{\partial E_{s_i}}{\partial p_j} \leq 0$, i.e., $\frac{\partial E_{s_i}}{\partial p_i} \leq 0$ since we know that $\frac{\partial E_{s_i}}{\partial p_j} = 0, s_j \in [L_s|s_i]$. *Property 5.2* already has proved $\frac{\partial E_{s_i}}{\partial p_i} \leq 0$.

On the other hand, for Case 1, if any source node s_i satisfies the source rejection criteria at the beginning, s_i is not rejected by Network in the final outcome of the game. Since optimal power $E^*_{s_i}(p_i)$ is a function of only source s_i's price, it does not get affected if other sources increment their prices. *Property 5.3* says that $\frac{\partial U_{s_i}}{\partial p_i} \geq 0$. Assume that $\bar{p_i}$ is the price for which source s_i obtains 0 power. Hence, when source s_i increments its price from c_i to some price, say p_i^{new}, this new price must be less than $\bar{p_i}$. This is because, in order to satisfy *Property 5.3*, it will ask such price which will increase its utility instead of obtaining 0 utility (achieved at price $\bar{p_i}$). However, for Cases 2 and 3, since Network assigns power among the sources in a greedy manner based on some metric, there is possibility that some source(s) get(s) rejected in the final outcome of the game due to the total power constraint.

Theorem 5.1: $\{E^*_{s_i}\}_{s_i \in L_s}$ and $\{p^*_i\}_{s_i \in L_s}$ are the SE for the source power allocation game.

Proof: Having obtained the prices $\{p^*_i\}_{s_i \in L_s}$ from the sources, due to *Property 5.1*, $U^s_n\left(\{E^*_{s_i}(p^*_i)\}_{s_i \in L_s}\right) \geq U^s_n\left(\{E_{s_i}(p^*_i)\}_{s_i \in L_s}\right)$. That implies, $\{E^*_{s_i}(p^*_i)\}_{s_i \in L_s}$ is the optimal response strategy for Network and the SE of the game. When the optimal response is released by Network, source s_i keeps increasing its price p_i until it reaches to p^*_i. According to *Property 5.3*, $U_{s_i}(p^*_i, E^*_{s_i}(p^*_i)) \geq U_{s_i}(p_i, E^*_{s_i}(p_i))$. Therefore, p^*_i is the optimal response for source s_i and the SE of the game.

Iterative Source Price Update Function

The sources increase their utilities by incrementing their prices from acceptable lower quantity, c_i (cost of power for delivering data) towards the optimal ones. The price update function of the sources can be designed as follows. In each iteration of the price update procedure until the convergence happens, for the sources $s_i \in L^1_s$ in Case 1, it applies that

$$\frac{\partial U_{s_i}}{\partial p_i} \;=\; \frac{\partial}{\partial p_i}\left[(p_i - c_i)E^*_{s_i}(p_i)\right] \;=\; E^*_{s_i}(p_i) + (p_i - c_i)\frac{\partial E^*_{s_i}(p_i)}{\partial p_i} \geq 0.$$

By *Property 5.2*, we know that $\frac{\partial E^*_{s_i}(p_i)}{\partial p_i} < 0$. After re-arranging, the above equation appears to

$$p_i \leq c_i - E^*_{s_i}(p_i)\left[\frac{\partial E^*_{s_i}(p_i)}{\partial p_i}\right]^{-1}. \tag{5.17}$$

Here, it is important to note that the value of $\frac{\partial E^*_{s_i}(p_i)}{\partial p_i}$ is negative before p_i rises to p_i^*. For the sake of simplicity, the price update procedure can be represented in vector form, $\mathbf{p} \leq \mathbf{I}(\mathbf{p})$, where $\mathbf{p} = \{p_i\}_{s_i \in L_s^1}$; $\mathbf{I}(\mathbf{p}) = \{I_i(\mathbf{p})\}_{s_i \in L_s^1}$, where $I_i(\mathbf{p})$ is the price update function for source s_i. Consequently, each iteration of the price update algorithm can be expressed as $\mathbf{p}(t+1) = \mathbf{I}(\mathbf{p}(t))$. $\mathbf{I}(\mathbf{p})$ is a standard function, and it has the similar properties as a standard function has. Because of these properties, the authors in [104] have proved that starting from some initial power vector $\mathbf{p}$, after n iterations, $\mathbf{I}^n(\mathbf{p})$ produces unique fixed prices. The properties of the standard function $\mathbf{I}(\mathbf{p})$ have been proved in *Appendix A*. For Case 2, $E^*_{s_i}(p_i), \forall s_i \in L_s^2$ is non-differentiable with respect to p_i. Therefore, the price update function for this case is $\mathbf{p}(t+1) = \mathbf{p}(t) + \boldsymbol{\delta}$. Whereas, for Case 3, the sources $\forall s_i \in L_s^1$ follow the similar price update procedure as Case 1, and the sources $\forall s_i \in L_s^2$ follow the procedure in Case 2.

5.3.2 Game Theoretical Relay Power Allocation

In order to enhance the performance of the network further, we present the formulation of another Stackelberg game in order to distribute relay power E'_r among the sources in set L_s^1 below.

1. Sources/Buyers: We first model the sources as followers who aim to get most benefits at the least possible payment. The utility function of source $s_i, s_i \in L_s^1$ can be defined by

$$U_{r_i} = \eta \left(\frac{E^*_{s_i} G_{s_i,d}}{\sigma^2} + \frac{E^*_{s_i} G_{s_i,r} E_{r_i} G_{r,d}}{\sigma^2 \left(E^*_{s_i} G_{s_i,r} + E_{r_i} G_{r,d} + \sigma^2 \right)} \right) - \lambda E_{r_i}. \tag{5.18}$$

   η is the gain per unit of SNR, and λ denotes the price per unit of power sold by the centralized node "Network". E_{r_i} can be explained as the amount of relay power source s_i would like to buy from Network when the price λ is announced.

2. Network/Seller: Network is modeled as a leader who aims to maximize its revenue by setting a proper price. Constant c is introduced to denote the cost per unit of power. The utility function of Network is defined as

$$U_n^r = (\lambda - c) \sum_{s_i \in L_s^1} E_{r_i}^*(\lambda). \tag{5.19}$$

λ has the same meaning as in Equation 5.18. In order to earn profit, the price must be higher than the cost, which means $\lambda > c$. $E_{r_i}^*(\lambda), s_i \in L_s^1$ depends not only on source s_i's channel condition, but also depends on the global price. If Network asks a larger price than what the source expects, the source will buy less power. On the other hand, if the price is too low, the profit obtained by Equation 5.19 will be unnecessarily small. So, there is a tradeoff for setting the price. A proper price can distribute the total allowable relay power among the sources optimally.

Analysis of Relay Power Allocation Game

In this section, we first show that given a price λ, there exists an unique SE of each source game. Then based on this, we prove, there is a unique optimal equilibrium for the whole Stackelberg game. Moreover, we design a price update function for Network and prove its convergence to the unique equilibrium in the following subsection.

(i) Source Level Analysis: Each source node ($s_i \in L_s^1$) determines how much power it should buy to maximize its utility. According to Equation 5.18, the power source s_i will buy under the price λ can be determined by solving the equation $\frac{\partial U_{r_i}}{\partial E_{r_i}} = 0$.

$$E_{r_i}^*(\lambda) = \begin{cases} 0 & \text{if } \lambda \geq \lambda_{r_i}^{ub} \\ \frac{1}{G_{r,d}}\left[\sqrt{\frac{\eta B_{r_i} E_{s_i}^* G_{s_i,r} G_{r,d}}{\lambda\sigma^2}} - B_{r_i}\right] & \text{if } \lambda_{r_i}^{ub} < \lambda < \lambda_{r_i}^{lb} \\ E_r' & \text{if } \lambda \leq \lambda_{r_i}^{lb} \end{cases}, \tag{5.20}$$

where $B_{r_i} = E_{s_i}^* G_{s_i,r} + \sigma^2$, $\lambda_{r_i}^{lb} = \frac{\eta B_{r_i} E_{s_i}^* G_{s_i,r} G_{r,d}}{\sigma^2 (E_r' G_{r,d} + B_{r_i})^2}$ and $\lambda_{r_i}^{ub} = \frac{\eta E_{s_i}^* G_{s_i,r} G_{r,d}}{B_{r_i}\sigma^2}$.

(ii) Network Level Analysis: Network needs to find a global optimal price so as to maximize its revenue. Given $E_{r_i}^*(\lambda), s_i{\in}L_s^1$, the best price can be obtained by solving the following equation

$$\frac{\partial U_n^r}{\partial \lambda} = \sum_{s_i \in L_s^1} E_{r_i}^*(\lambda) + (\lambda - c) \sum_{s_i \in L_s^1} \frac{\partial E_{r_i}^*(\lambda)}{\partial \lambda}. \tag{5.21}$$

Hence, the best price is given by

$$\lambda^* = f\left(\{G_{s_i,r}\}_{s_i \in L_s^1}, G_{r,d}, c, \eta, \sigma^2\right). \tag{5.22}$$

Lemma 5.2: The optimal price is inside the interval $[\lambda_{lb},\ \lambda_{ub})$, where $\lambda_{lb} = \max\limits_{s_i \in L_s^1} \lambda_{r_i}^{lb}$ and $\lambda_{ub} = \max\limits_{s_i \in L_s^1} \lambda_{r_i}^{ub}$.

Proof: We prove the lower bound by contradiction. If $\lambda < \max_{s_i \in L_s^1} \lambda_{r_i}^{lb}$, one source will definitely obtain E_r' relay power and at most one source will obtain some relay power which results in the amount of allocated relay power is more than the allowable power E_r'. This is not a feasible solution. On the other hand, if $\lambda = \max_{s_i \in L_s^1} \lambda_{r_i}^{lb}$, one source will obtain E_r' relay power and the rest others may obtain 0 relay power, which is considered as a feasible solution. Considering the feasibility of the allocated relay power among all sources, λ should be $\geq \max_{s_i \in L_s^1} \lambda_{r_i}^{lb}$.

For the upper bound, if $\lambda \geq \max_{s_i \in L_s^1} \lambda_{r_i}^{ub}$, all sources obtain 0 relay power which is also not an expected solution. Therefore, λ must be $< \max_{s_i \in L_s^1} \lambda_{r_i}^{ub}$ in order to assign some relay power among the sources.

Properties of Relay Power Allocation Game

In this subsection, we elaborate the properties of the relay power allocation game and prove that the solutions derived in Equations 5.20 and 5.22 are the unique optimal equilibrium for each round of this game.

Property 5.4: Given price λ, U_{r_i} is concave with respect to E_{r_i}, when $E_{r_i} > 0$ and $E_{r_j}, \forall s_j \in [L_s^1|s_i]$ are fixed quantity.

Proof: Taking the second order derivative of U_{r_i} in Equation 5.18, we get

$$\frac{\partial^2 U_{r_i}}{\partial E_{r_i}^2} = \frac{-2\eta B_{r_i} E_{s_i}^* G_{s_i,r} G_{r,d}}{\sigma^2 (B_{r_i} + E_{r_i} G_{r,d})^3} < 0. \tag{5.23}$$

Moreover, U_{r_i} is continuous in E_{r_i}. So, when $E_{r_i} > 0$, U_{r_i} is concave with respect to E_{r_i}.

Property 5.5: The best amount of power bought by each source decreases as the price increases.

Proof: Taking the derivative of $E_{r_i}^*(\lambda)$ with respect to λ, we obtain

$$\frac{\partial E_{r_i}^*(\lambda)}{\partial \lambda} = \frac{-1}{2G_{r,d}} \sqrt{\frac{\eta B_{r_i} E_{s_i}^* G_{s_i,r} G_{r,d}}{\sigma^2 \lambda^3}} < 0. \tag{5.24}$$

This justifies the decrementing trend of the optimal relay power for sources $s_i \in L_s^1$ with the incrementing λ.

Property 5.6: If $\lambda > c$, U_n^r is concave with respect to price λ and the power consumption is the optimized purchase amount.

Proof: From Equations 5.21 and 5.24, we obtain

$$\frac{\partial^2 U_n^r}{\partial \lambda^2} = 2 \sum_{s_i \in L_s^1} \frac{\partial E_{r_i}^*(\lambda)}{\partial \lambda} + (\lambda - c) \sum_{s_i \in L_s^1} \frac{\partial^2 E_{r_i}^*(\lambda)}{\partial \lambda^2}, \tag{5.25}$$

$$\frac{\partial^2 E_{r_i}^*(\lambda)}{\partial \lambda^2} = \frac{3}{4G_{r,d}} \sqrt{\frac{\eta B_{r_i} E_{s_i}^* G_{s_i,r} G_{r,d}}{\sigma^2 \lambda^5}}. \tag{5.26}$$

Since $\sqrt{\frac{1}{\lambda^3}} > \sqrt{\frac{1}{\frac{\lambda^5}{(\lambda-c)^2}}}$, $\frac{\partial^2 U_n^r}{\partial \lambda^2} < 0$, which justifies the concavity property of U_n^r with respect to λ.

Property 5.7: Given the relay power price λ, $E_{r_i}^*(\lambda)$ is a non-decreasing function of $E_{s_i}^*$ and $G_{s_i,r}$.

Proof: See *Lemma 5.2* of [102].

Theorem 5.2: $\{\{E_{r_i}^*\}_{s_i \in L_s^1}, \lambda^*\}$ solved in above discussions are the SE for the proposed game.

Proof: According to *Property 5.4*, given λ and the power of other sources $\{E_{r_j}^*\}_{s_j \in [L_s^1 | s_i]}$ is constant, the best response of source s_i is $E_{r_i}^*(\lambda)$. And, according to *Property 5.6*, incrementing λ gradually increases the value of U_n^r until it reaches λ^*. It implies that

$$U_{r_i}(E_{r_i}^*(\lambda^*)) \geq U_{r_i}(E_{r_i}(\lambda^*)),$$

$$U_n^r(\{E_{r_i}^*(\lambda^*)\}_{s_i \in L_s^1}) \geq U_n^r(\{E_{r_i}^*(\lambda)\}_{s_i \in L_s^1}).$$

So, $\{\{E_{r_i}^*\}_{s_i \in L_s^1}, \lambda^*\}$ are the SE of the relay power allocation Stackelberg game.

Iterative Relay Power Price Update Function

In order to achieve appropriate convergence of this game, Network needs to update its price correctly in each round. According to *Property 5.6*, U_n^r is a concave function with respect to λ. Therefore, Network increases the price λ from its initial price until the convergence happens. For the sake of obtaining non-negative utility, Network sets the initial price c. According to Equation 5.21 and *Property 5.5*, if $\frac{\partial}{\partial\lambda}U_n^r > 0$, we have

$$\lambda < c - \left(\sum_{s_i \in L_s^1} E_{r_i}^*(\lambda)\right)\left(\sum_{s_i \in L_s^1} \frac{\partial E_{r_i}^*(\lambda)}{\partial\lambda}\right)^{-1}.$$

We denote $I(\lambda)$ as

$$I(\lambda) = c - \left(\sum_{s_i \in L_s^1} E_{r_i}^*(\lambda)\right)\left(\sum_{s_i \in L_s^1} \frac{\partial E_{r_i}^*(\lambda)}{\partial\lambda}\right)^{-1}. \tag{5.27}$$

Hence, the price update method is $\lambda(t+1) = I(\lambda(t))$. $I(\lambda)$ fulfills the properties of a standard function which has been proved in *Appendix B*. Setting the initial price as c (i.e.,λ_{lb}), λ will converge to a unique equilibrium after sufficient iterations.

5.3.3 Further Discussion

Next, we will briefly discuss the possible implementation of the proposed game theoretical solution for this power allocation problem. As noted before, there is a centralized entity called "Network" in the network. Network is responsible to lead two games in order to obtain the transmit and relay power allocation of the sources. There is a total power constraint in the network which is the sum of required power for both transmit and relay operations of the sources. In order to decide appropriate convergence of two games separately, Network needs to know how much power is separately allocated for both games. As discussed before, in order to decide the amount power for both games, Network should be aware of the complete CSI of the network. On the assumption of using block fading channel model, at the beginning of the first time slot (the entire transmit operation requires two phases or time slots), a training process is conducted for Network to obtain global CSI. Research on the efficient channel training and estimation can be found in [105, 106]. The training and estimation is performed at the relay and the destination in order to obtain the channel gains from the sources to themselves. For collecting the channel gain from the relay to the destination, another round of training and estimation is performed at the relay. Finally, all these acquired channel gains are notified to Network by the feedback method. Network can be a separate computationally powerful entity in the network or; any of the sources or the relay can play this role depending on their computational capability.

Once Network knows the allowable power for two games, first it starts the game with the sources for their transmit power. When the convergence is achieved for this game and set L_s^1 is not empty, Network initiates the second game, and continues until the convergence happens. In order to improve the performance of the network further, Network can re-distribute the power especially for the case when both sets L_s^1 and L_s^2 are non-empty. The way how Network does this is described as follows.

1. Sort the list L_s^2 in the descending order of $G_{s_i,d}, s_i{\in}L_s^2$, and denote the sorted list as OL_s^2.

2. Sort the list L_s^1 in the ascending order ot $G_{s_i,d}, s_i{\in}L_S^1$, and denote the sorted list as OL_s^1.

3. Take source s_i from OL_s^2 which is not assigned with full power E_s^{max}.

4. Take source s_j from OL_s^1.

5. Compute $x = \min\left(E_{s_j}^*, E_{r_j}^*\right)$ and $y = \min\left(2x, E_{s_i}^*\right)$. Add y to $E_{s_i}^*$, and subtract $y/2$ from $E_{s_j}^*$ and $E_{r_j}^*$.

6. If $E_{s_i}^* = E_s^{max}$, move s_i to the next source of list OL_s^2.

7. If $E_{s_j}^* = 0$ and $E_{r_j}^* > 0$, add $E_{r_j}^*$ to $E_{s_j}^*$.

8. If $E_{s_j}^* = 0$, move s_j to the next source of list OL_s^1.

9. If s_i or s_j does not point to a valid source, terminate this process, otherwise repeat the steps from 5.

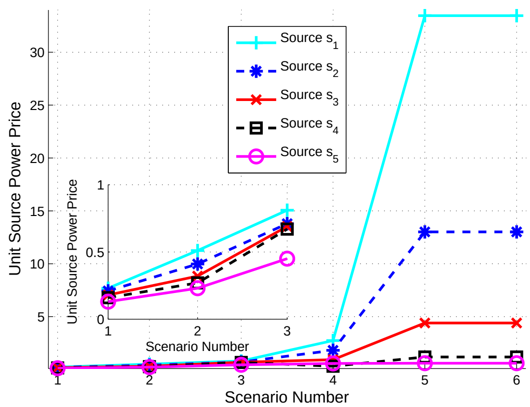
With this implementation, we actually assume that Network is trustworthy. All sources believe that Network will not change the parameter values (e.g., CSI), however conducts all these operations in the systematic manner.

5.4 Simulation Results

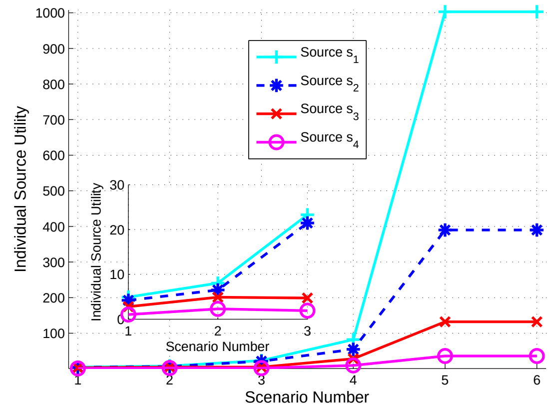
In order to evaluate the game theoretical solution, we undertake the similar network setup, channel model and the parameters as given in Section 3.3.4 of Chapter 3. Similar simulation scenario is undertaken as was taken for evaluating the performance of the centralized approach. For the game theoretical solution, there are some parameters, which we set as $a = 100$, $\eta = 100$[12].

Within a certain price range, the utility of each source is concave. Ideally, convergence of the source power allocation game should be decided by the individual source

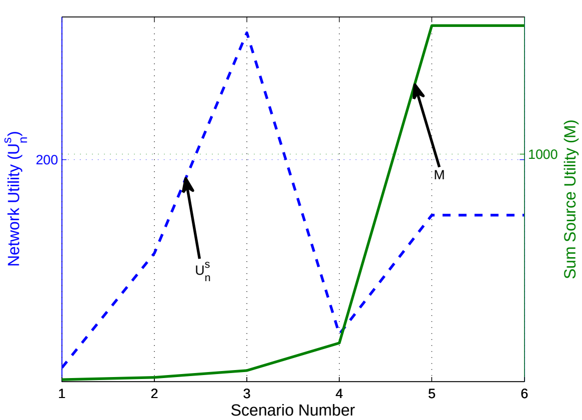
[12]Setting a and η as arbitrarily random values, only the resultant converged prices $\{p_i\}_{s_i \in L_s}$ and λ are affected. However, main outcome of the game $\{E_{s_i}\}_{s_i \in L_s}$ and $\{E_{r_i}\}_{s_i \in L_s^1}$ are unique regardless of the values of a and η.



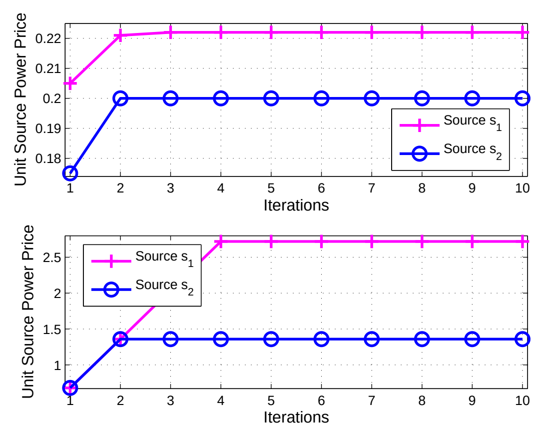
(a) Converged price comparison among all sources.



(b) Individual converged utility comparison among all sources.



(c) Converged Network utility and sum source utility comparison.



(d) Convergence for scenario 1 (up) and scenario 4 (down).

Figure 5.2: Source power allocation game.

nodes. However, to make the game theoretical solution comparable with the centralized one, Network should be aware of total power constraint, i.e., the sum power sold by the sources should be equal to the total allowable source power. Therefore, if the sum of converged power is less than the total source power budget, the remaining power is distributed among the sources based on their contribution towards the utility of Network node. On the other hand, if the sum of the converged power is less than the total power budget, overflown power is subtracted from the sources based on their inverse utility contribution.

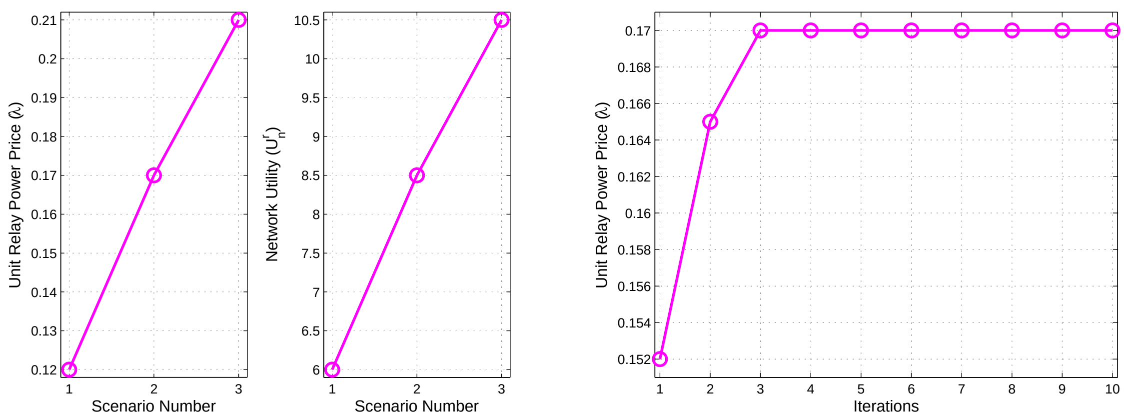
Figures 5.2(a), 5.2(b) and 5.2(c) compare individual converged source power price p_i, utility U_{s_i}, sum source utility M, and Network node's utility U_n^s for different scenarios explained previously. As the sources move towards the relay or the destination, prices asked by them are increased. Since their channel condition gradually improves with the increasing scenario number, they demand high price for unit source power. For scenario number 1, 2 and 3, the price of the sources is the sum of relayed and direct link contribution towards maximizing Network's utility. Whereas for other scenarios, the price is only for the direct link contribution. Since the absolute channel condition of scenarios 5 and 6 is almost same, the prices of the sources are almost equivalent for these two scenarios. In each scenario, the sources with better channel condition contribute more towards maximizing Networks's utility, and hence they demand higher price compared to others. Furthermore, the utility of individual source follows the same trend as its price since the utility is the product of its price and sold power. At the convergence point, all scenarios satisfy the same power constraint, and hence the utility of individual source is increased with the increasing price. Consequently, the sum utility of all sources M is increased with the increasing scenario number. In Figure 5.2(b), we have skipped the utility of source s_5 because of its achieved 0 utility resulting from nearly 0 sold power. Since the channel condition of the sources is getting better with the increasing scenario number, even though M is subtracted form the first factor of Network's utility, because of using moderately larger value of a, the utility of Network is increased as well. There is a discrepancy between scenario 3 and scenario 4 in terms of the utility of Network node, this is because former one uses the relay for transmission, whereas the latter one does not. Moreover, the converged price of each source for scenario 4 is close to the upper bound of its price, and hence the value of M does not deviate much from the first factor of U_n^s. However, the scenarios without relay power follow the same increasing trend with the better channel quality condition because of the increasing first factor of U_n^s compared to its M. Table 5.1 shows the final outcome of the game, a detailed breakdown of the source power. In order to have comparison, the table also shows the power obtained from the centralized optimal solution.

At the beginning, when the game is started, Network selects the set L_s of beneficiary sources following the procedure in *Remark 5.1.* Taking the consideration of their own cost, the sources announce their prices. Consequently, Network calculates the amount of power it wants to buy from the sources for maximizing its own utility. Considering this interaction as one iteration, Figure 5.2(d) shows the convergence process for scenarios 1 and 4. For scenarios 4, 5 and 6, convergence speed depends on the step size $\delta_i, s_i \in L_s$. The larger the step size, the speedier the convergence.

Similar to the source power allocation game, for the relay power allocation game, the convergence point should be decided by the seller of the game, i.e., Network. In order to have valid comparison with the centralized solution, the sum relay power should meet some power constraint. At the convergence point of Network node, the power constraint should be satisfied. If the constraint is not satisfied at the

Table 5.1: Source transmit power comparison between optimal and game theoretical solutions.

Scenario Number	s_1		s_2		s_3		s_4		s_5	
	Opt	Game	Opt	Game	Opt	Game	Opt	Game	Opt	Game
1	28.29	24.32	26.46	19	15.24	16	0	10.68	0	51e-3
2	29.12	25.1	28.21	21.0	12.66	12.13	0	7.1	0	4e-3
3	30.0	27.23	30	24.4	10	11.55	0	6.80	0	71e-4
4	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	0	0
5	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	0	0



(a) Relay power price (left) and converged Network utility (right) comparison.

(b) Convergence for scenario 2.

Figure 5.3: Relay power allocation game.

convergence point, relay power of each source is adjusted based on its contribution towards the utility of Network node.

Figure 5.3(a) shows the increasing price and utility of Network node for the relay power allocation game with respect to different scenarios. Similar to the source power allocation game, the increasing scenario number implies better channel quality condition. It means, unit relay power results in more contribution towards maximizing the utility of individual source. Hence, Network demands larger relay power price while selling its power to the sources. Since the power budget is same for all scenarios, with the increasing price, the utility of Network node has increasing trend as well. Table 5.2 shows the final outcome of the game. It compares the power obtained by the relay power allocation game with the centralized optimal solution. It seems, individual source who has better channel condition buys more power compared to others. This is because, given the unit price, the source with better channel condition incurs larger utility compared to those with worse ones. Moreover, the utility of individual source is the increasing function of its transmit power. Since the source with better channel condition is assigned larger transmit power from the previous game, this source is likely to be assigned with more relay power compared to others. Therefore, the larger demanded power is well adjusted balance between their utility and price.

After carefully considering the initial price, when Network announces it, the sources demand power which is the balance between their individual utility and price. Considering this interaction between the sources and Network as a single step, Figure 5.3(b) shows the convergence behavior of the relay power allocation game at scenario 2. The figure shows the increasing price of Network node with the increasing iteration until the convergence happens.

Table 5.2: Source relay power comparison between optimal and game theoretical solutions.

Scenario	s_1		s_2		s_3		s_4		s_5	
Number	Opt	Game	Opt	Game	Opt	Game	Opt	Game	Opt	Game
1	36.19	46	13.78	4.0	0	0	0	0	0	0
2	36.96	46.3	13.031	3.7	0	0	0	0	0	0
3	32.24	43.5	17.75	6.5	0	0	0	0	0	0

Taking the power presented in Table 5.1 and Table 5.2, Figure 5.4 compares the throughput obtained from the game theoretical solution with the centralized optimal solution.

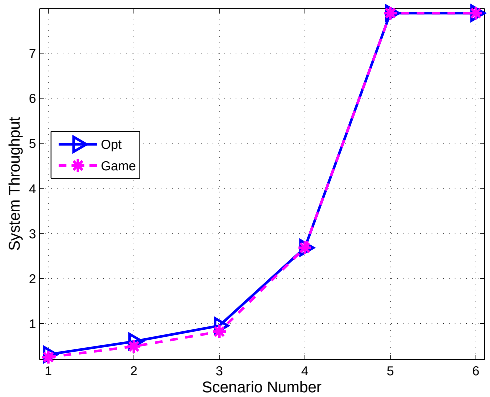


Figure 5.4: System throughput comparison between optimal and game theoretical solutions.

5.5 Summary of the Chapter

In this chapter, we have proposed a game theoretical solution of the problem described in Chapter 3. In the centralized solution, all nodes work selflessly towards maximizing the network capacity. The game theoretical solution proposed in this chapter can model the selfish behavior of the nodes in the network. Since joint single-source and single-relay power allocation is even a non-convex problem, two Stackelberg games are appropriate to determine the transmit and relay power of the sources. A centralized entity namely Network is required to make a bridge between these two games, and it is the leader of both games. For disseminating transmit power among the sources, Network plays the buyer level game and the sources are power sellers. Their roles are interchanged for disseminating relay power among the sources. Extensive numerical results have been provided in order to show different properties, convergence condition of the solution. Finally, we have shown that the game theoretical solution achieves comparable performance with the centralized optimal solution.

Chapter 6

Uplink Scheduling and Resource Allocation for Heterogeneous Traffic Networks

6.1 Introduction

Based on OFDM and OFDMA principals, LTE and LTE-Advanced permit granular subcarrier division among the diverse users into time-frequency-based resource units called RBs. Despite numerous advantages of OFDM and OFDMA, their major disadvantage is, their waveforms have high PAPR. To reduce PAPR, LTE and LTE-Advanced agree on using SC-FDMA for the uplink transmission which imposes contiguous RB allocation to a user. Typical scheduling of such systems involves the determination of the set of users, assignment of RBs to these users and setting of transmit power for each RB. Uplink scheduling even for the conventional LTE systems with best effort traffic is considered challenging because of individual users' power constraints, the discrete nature of RB assignment while maintaining contiguity pattern. With the invention of numerous applications with distinct fascinating features, now a days, LTE systems are more likely to carry heterogeneous traffic with different QoS demands. Scheduling RBs while satisfying diverse QoS of different new applications brings more challenges to the uplink scheduling problem on the top of its own inherent difficulties. Besides all these system and traffic specific issues, the service providers may want to have domination on the RB allocation determined by their policy. This introduces further challenges to the uplink scheduling problem in LTE systems conveying heterogeneous traffic.

Resource allocation problems for OFDM-based networks especially downlink LTE systems have appeared in some survey papers recently [51], [52]. In homogeneous traffic networks, for two types of traffic, e.g., elastic traffic, traffic with QoS requirements, scheduling problems have extensively been studied. For the elastic traffic, in order to perform the scheduling decision, the utility of the users is represented by a concave function [107]. Subject to the objective of maximizing the sum of general concave utility functions, one recent work [108] has proposed some computationally efficient algorithms. Channel-aware throughput maximization technique for such systems is known as maximum throughput (MT). For fair resource allocation, typical proportional fair (PF) metric [109], [110], [111] is the ratio of the instantaneous data rate on a RB and the past average throughput. In [112], [113], PF scheme is formulated as an optimization problem with the objective of maximizing achieved throughput under the typical constraints of LTE systems.

Among QoS-aware schedulers in homogeneous traffic networks, [55] ensures guaranteed throughput by separating jobs in the time and frequency domains. In the time domain, they first separate traffic based on their current and pre-specified settled throughput. Then, they apply some PF scheme on the same priority users in order to obtain final RB assignment. More works on providing throughput guarantee for real time flows include [114], [115]. [116] calculates the priority indicator using head of line (HOL) packet delay, and ensures target delay bound for the delay sensitive traffic. Modified largest weighted delay first (M-LWDF)[117] has been introduced for OFDMA-based networks in [53] by putting past throughput in the denominator of the original metric. Two promising strategies, i.e., LOG and EXP rules have been explained in [54]. Their RB selection metrics are based on the logarithmic and exponential functions of the packet HOL delay, respectively. Unlike LOG, EXP scheduler takes the overall network status into account. Similar to these, there is another scheduling rule, i.e., maxweight (MW) [118] for delay sensitive traffic. Analytical results in terms of queue distribution for the schedulers with these rules are given in [119], [120], [121]. In [122], the authors adapt EXP rule by taking the characteristics of PF and exponential function of the packet HOL delay into account. A two level packet scheduler working in the larger LTE frame, and then in the granular frame level for real time traffic has been given in [123]. A similar approach, however working in 3 discrete levels has been presented in [124]. The third level of this scheduler is called cut-in process which discards packets whose delay deadline has been expired. A recent work [125] combines EXP rule, cooperative game and virtual token mechanism in order to ensure bounded delay and guaranteed bit rate for users.

Scheduling problem in heterogeneous traffic networks is considered more challenging compared to the homogeneous one because of the conflicting requirements of different traffic. There are three design issues that need to be considered while formulating the scheduling problem of an ideal scheduler for heterogeneous traffic networks. Wireless users experience varying channel quality condition time to time due to the stochastic fading effects, and hence their achievable data rates are affected. The scheduler should choose the user with good channel quality condition in order to maximize the system throughput. However, if the user with better channel quality condition is always selected, the users with worse channel condition may starve. This issue is known as fairness, and another very important factor to keep into consideration. Third, different applications have different QoS requirements and the utility function of the scheduler should be defined in such a way that it has the ability to capture those metrics simultaneously and effectively.

Similar to homogeneous traffic networks, the downlink scheduling has extensively been studied for OFDM-based networks carrying heterogeneous traffic. These works mainly adopted three categories of designs. First, some works gave strict priority to the high priority traffic compared to the low priority ones, such as [57, 58]. [57] has proposed a solution for VoIP and data traffic while giving strict priority to VoIP users. A scheduling policy with strict priority across the classes is also studied in [58]. Within a class, the proposed scheduler does chunk by chunk resource allocation. The work in [126] has given strict priority to SIP traffic over other data traffic. Second, the authors in [59] have addressed mixed traffic (delay sensitive, guaranteed throughput, best effort) by designing the scheduler on the time and frequency domains. The priority sets are populated based on the QCI of each data flow, and classified as GBR and non-GBR. Third, the scheduler is designed by representing the satisfaction of users by the utility function. In [60, 61], different utility functions are used to represent different types of traffic (traffic with data rate requirement, traffic with delay constraint). Although [60] did not consider the channel condition of users in the scheduling decision, [61] took all required issues into account. Besides these, there are some general mechanisms for handling mixed traffic. For example, the authors in [54] have proposed a low complexity RB allocation algorithm using LOG/EXP, EDF rules, and tested their scheme with mixed streaming and live video traffic. The work in [127] studied a scheduling policy that gives equal priority to all packets with different QoS unless their delay reaches close to their deadline.

Unlike downlink operation, not much works have been conducted for the uplink LTE systems. Most works focused on throughput maximization objective. For example, [10, 128, 129, 11] have given different heuristics based on different principles upon stating that the problem is NP-hard. Although these works take RB contiguity constraint into account, skip individual users' power constraint. However, this constraint is an important factor in scheduling decision, and limits the overall system performance. On the other hand, there are few works for the uplink scheduling with QoS assurance. These works are mainly for homogeneous delay sensitive traffic [62, 63]. One recent work on energy efficient QoS assured scheduler is [64], where QoS is considered user's instantaneous rate. For heterogeneous traffic environments, recently, [65] has proposed a scheduler which is based on the time and frequency domain scheduler [59]. One drawback of this work is, they did not consider the traffic class in their design. Moreover, they evaluated the performance of this scheduler in their customized simulation scenario, which is not practical.

Having noticed the drawback of [65], and in order to design a utility based scheduler for the heterogeneous traffic environment, in this chapter, we have proposed an uplink scheduling scheme for a LTE system carrying heterogeneous traffic. This work is based on the utility function [130] has used. The difference is, their solution is suitable for the downlink scheduling, and designed for CDMA-based systems. The utility function of [130] is general, one single function can handle diverse types of traffic, instantaneous channel condition and so forth. Typically, this utility function is used in economics in order to maintain social welfare (another name of fairness) of the society, however rarely applied in communications. Moreover, this utility function can distinguish inter or intra class based traffic prioritization. Besides capturing three key design issues of the heterogeneous traffic environment by this utility function, in order to assist the service providers, the proposed uplink scheduling problem consists of a constraint which imposes some control on RB allocation. The distinctive novelty and contributions of this chapter are briefly summarized as follows.

1. For representing the satisfaction of a user, we have adopted the utility function already used for the downlink scheduling operation of a CDMA-based system [130]. Although we have adopted their utility function and same definition of QoS measures for different traffic, we have solved a distinct problem for a different system, i.e., uplink scheduling of LTE systems. Furthermore, the way they have solved the downlink scheduling problem is different compared to us. Given pre-defined physical layer data rate of users and given certain system capacity, their work schedules a set of users instead of CDMA tones. Whereas, our scheduling scheme determines every single RB-user mapping, their assigned power which is the true notion of an ideal scheduler given the incoming packets in each scheduling epoch.

2. For obeying every detailed specifications of the LTE standard, our formulated problem takes individual user's power and RB contiguity constraints into account. Individual user's power constraint is required in order to correctly measure the overall system performance. On the other hand, in order to avoid high PAPR of the generated waveform, we assign contiguous RBs to each user.

3. Besides the standard specific constraints, our formulation consists of another factor, opportunity cost function. Physical meaning of this function is how much utilization of resource the service provider can sacrifice in order to achieve other system or user specific benefits. The cost function can work in both granular and aggregate resource utilization levels. For this work, we assume it

to be dependent on RB utilization as the latter one can be achieved through the former one, however not vice versa. Furthermore, service providers may relate opportunity cost function to revenue as desired.

4. By setting certain parameter for the opportunity cost function, our scheduling scheme can be transformed to two extreme ends of scheduling mechanisms, i.e., MT and PF schedulers. By providing enough evidence, we have proved this statement analytically.

5. Dual decomposition method has been used in order to show the optimal solution structure of our formulated problem. Finally, from the guiding principles of the optimal solution, we have given a low complexity scheduling algorithm.

6. Extensive simulation has been conducted assuming the network has best effort traffic, traffic with throughput requirement and traffic with delay bound. While comparing with other scheduling solutions of LTE systems, we have proved the effectiveness and efficacy of our scheme.

The rest of the chapter is organized as follows, Section 6.2 gives the overview and components of the system while elaborating detailed formulation of the problem. Solution approach and the resultant algorithm are presented in Section 6.3. We investigate some interesting characteristics of our scheduling scheme in Section 6.4. In order to show the effectiveness of the proposed scheme, we provide simulation results followed by the simulation methodology in Section 6.5. Finally, Section 6.6 concludes the chapter.

6.2 System Model and Problem Formulation

We consider an uplink scheduling problem of a typical LTE cellular network. We denote the set which contains available RBs at each scheduling epoch by $\mathbf{N}$. The number of users in the system is M. The users are again categorized into C classes, where class c has higher priority than class $c+1$. Let M_c denote the number of class c users; $M = \sum_{c}^{C} M_c$ and the set holding all users is $\mathbf{M}$. Each class is accompanied with diverse kinds of QoS criteria. At each scheduling instant, all users transmit their CSI to the base station. In the channel coherent time, based on the traffic type, channel state and their QoS requirements, the base station assigns appropriate RBs to the corresponding users as well as determines power of each RB. The purpose of the scheduler equipped in the centralized controller is to maximize the sum satisfaction of all users.

6.2.1 Problem Formulation

Since SC-FDMA-based resource allocation allows service providers to allocate resource in granular RB level, for a particular user i of a certain class c, we define its utility function for each and every RB $j, \forall j \in \mathbf{N}$. At scheduling time instant t, the satisfaction of user i of class c on RB j is perceived by the utility function $U_{cij}\left(\{X_{cij}^{z}(t)\}_{z=1}^{m_{ci}}\right)$, where $\{X_{cij}^{z}(t)\}_{z=1}^{m_{ci}} = \{X_{cij}^{1}(t),\ X_{cij}^{2}(t),\ \cdots\ X_{cij}^{m_{ci}}(t)\}$. $\{X_{cij}^{1}(t),\ \cdots\ X_{cij}^{m_{ci}}(t)\}$ are computed quantitative QoS measures in terms of the (c,i)th user's satisfactions in LTE uplink systems during the scheduling decision of RB j at time instant t. Typically, QoS measures are the average throughput, current data rate, average delay, etc. m_{ci} represents the maximum number of QoS measures for user (c,i). Therefore, we can write the objective function as

$$\max \sum_{c}^{C} \sum_{i}^{M_c} \sum_{j}^{N} U_{cij}(\{X_{cij}^{z}\}_{z=1}^{m_{ci}}(t)). \tag{6.1}$$

If x_{cij} denotes the fraction of RB j allocated to user (c, i), the total allocation across all users should be no larger than 1, i.e.,

$$\sum_{(c,i)\in\mathbf{M}} x_{cij}(t) \leq 1, \ \forall j\in\mathbf{N}. \tag{6.2}$$

According to the LTE standard, there is a constraint on x_{cij} to be an integer $\{0, 1\}$. We will see later, how we have handled this constraint while solving this problem. Besides this, due to the contiguity constraint of SC-FDMA, RBs are allocated to user (c, i) in a contiguous manner. It implies

$$\begin{aligned} &\text{if } x_{cin}(t) = 1 \text{ \&\& } x_{ci(n+1)}(t) = 0, x_{cij}(t) = 0, \ n+2\leq j\leq N, \\ &\text{if } x_{cin}(t) = 1 \text{ \&\& } x_{ci(n-1)}(t) = 0, x_{cij}(t) = 0, \ 1\leq j\leq n-2. \end{aligned} \tag{6.3}$$

Each user (c, i) has power limitation during a scheduling epoch, and its maximum power is denoted by P_{ci}. For transmitting on RB j, we denote the (c, i)th user's transmission power is p_{cij} and total power used on all allocated RBs cannot exceed the maximum power P_{ci}.[13]

$$\sum_{j\in\mathbf{N}} p_{cij}(t) \leq P_{ci}, \ \forall (c, i)\in\mathbf{M}. \tag{6.4}$$

Moreover, there are upper and lower bounds of each QoS measure, and these are pre-defined values set by the users. For example, the zth QoS measure of user (c, i)

[13] According to Figure 3 of [3], different levels of power for the allocated RBs of a user is allowed.

has an upper bound $\nu_{ci}^{z,max}$ and a lower bound $\nu_{ci}^{z,min}$ (e.g., maximum and minimum average throughput).

$$\nu_{ci}^{z,min} \leq \; X_{cij}^{z}(t) \leq \; \nu_{ci}^{z,max}, \;\; \forall (c,i) \in \mathbf{M}, \; \forall z, 1 \leq z \leq m_{ci}. \tag{6.5}$$

The constraints presented in Equations 6.2, 6.3, 6.4 and 6.5 are either enforced by the LTE standard or set by the users. Another constraint we would like to introduce is for the sake of service providers, i.e., opportunity cost. The concept of opportunity cost can be used to manage the tradeoff between fairness and resource utilization. In order to ensure fairness across the network, the scheduler may be forced to serve low rate generating users resulting in rate loss. To limit this rate loss, we have proposed the cost function called opportunity cost. Similar to other metrics, the opportunity cost is also defined in granular user and RB level. At scheduling instant t, while allocating RB j, we define the opportunity cost for user (c,i) as $OC_{cij}(t)$. $OC_{cij}(t)$ is a function of the rate for user (c,i) on RB j, $R_{cij}(t)$. $R_{cij}(t)$ is given by

$$R_{cij}(t) = x_{cij}(t)\log\left(1 + \frac{p_{cij}(t)e_{cij}(t)}{x_{cij}(t)}\right), \tag{6.6}$$

where e_{cij} is the normalized received SNR per unit of transmit power of user (c,i) on RB j from the base station. At the beginning of scheduling instant t, the scheduler gathers complete information of this metric for all users and all RBs. The maximum rate that the scheduler can earn out of RB j at time t is given by

$$R_j^{max}(t) = \max_{(c,i)\in\mathbf{M}} R_{cij}(t), \; \forall j \in \mathbf{N}. \tag{6.7}$$

Using these metrics, $OC_{cij}(t)$ is defined as follows.

$$OC_{cij}(t) = R_j^{max}(t) - R_{cij}(t). \tag{6.8}$$

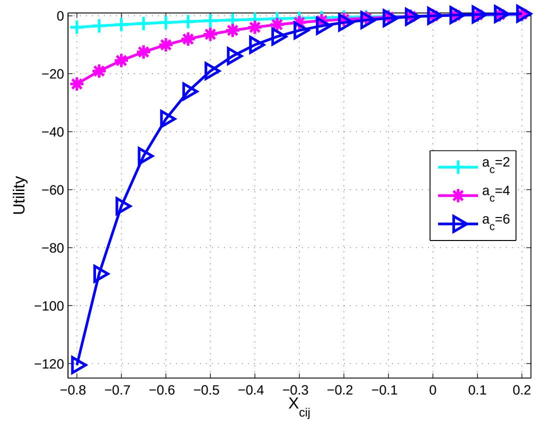
In other way, the opportunity cost is a measure of how much rate the network operator would forgo if user (c, i) is selected for the transmission on RB j at scheduling instant t, while there is some other user $(c, i)^*$ that generates the highest rate on this RB. Taking the interest of network operators into account, the objective function in Equation 6.1 can further be constrained by the opportunity cost function

$$OC_{cij}(t) \leq H,\ \forall (c, i) \in \mathbf{M}, \forall j \in \mathbf{N}. \tag{6.9}$$

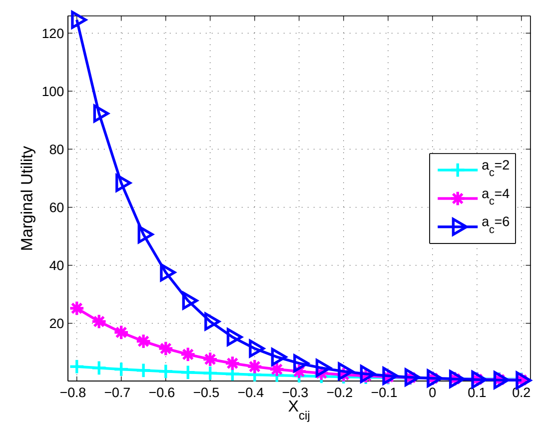
The network operator can determine the required level of fairness and resource utilization by choosing the appropriate value of H. If $H = \epsilon R_j^{max}$, the rate loss is no more than $\epsilon\%$ of the maximum achievable rate R_j^{max} on RB j. Moreover, $H = 0$ implies that the network operator cannot tolerate any rate loss, therefore it always picks the highest rate generating user while allocating any RB. In order to ignore the opportunity cost, the network operator can set $H = R_j^{max}$. In this case, all users are treated equally by the scheduler.

6.2.2 The Utility Function

In this section, we introduce a utility function which is able to meet all requirements of an ideal utility function. In order to satisfy all requirements of the problem described in the previous subsection, the feasible utility function should have the concavity property. The more the allocated resource, the more satisfied the user is, i.e., the utility function should be non-decreasing with $X_{cij}^z, \forall j \in \mathbf{N}, \forall z \in [1, m_{ci}]$. While allocating a RB at time instant t, if the values of all QoS measures for a user (c, i)



(a) Utility comparison for different a_c and X_{cij}.



(b) Marginal utility comparison for different a_c and X_{cij}.

Figure 6.1: Characteristics of the utility function.

are minimum, the utility value for that user attains its unique minimum value U_{ci}^{min}, whereas when the values reach to their maximum, the utility appears to its unique maximum value U_{ci}^{max}. In the rest of other cases, the utility value remains in between these two quantities, or in other way, $U_{ci}^{min} \leq U_{cij}(\{X_{cij}(t)\}_{z=1}^{m_{ci}}) \leq U_{ci}^{max}$. Furthermore, once the utility value for a user reaches to its maximum value, additional allocated resource cannot deviate it from its maximum quantity U_{ci}^{max}.

In addition of having above fundamental properties, the utility function should support inter-class prioritization. We denote a parameter $a_c, \forall c, 1 \leq c \leq C$ to distinguish inter-class prioritization. The larger the value of a_c, the higher the priority of class c. The proposed utility function is already used for the scheduling purpose in CDMA-based networks. As our problem is different, we interpret this function in a different way. The utility function for user (c, i) while allocating RB j at time instant t is given by

$$U_{cij}\left(\{X_{cij}^{z}\}_{z=1}^{m_{ci}}(t)\right) = 1 - e^{-a_c \sum_{z=1}^{m_{ci}} X_{cij}^{z}(t)}. \tag{6.10}$$

We plot the utility function with respect to arbitrary QoS measure X_{cij} for 3 users with different a_c in Figure 6.1(a). We observe, the utility function has the diminishing property. When the quantitative value of X_{cij} is low, the rate of change of the utility (slope) is larger. It implies, if the scheduler gives priority to the user with low QoS measure, the contribution of this user towards the maximization of overall utility is higher compared to others. Hence, the scheduler needs to take this into account while making the scheduling decision of RB j. In order to show the decreasing trend of the slope with the increasing value of X_{cij}, we plot Figure 6.1(b). We define the slope of the utility function as marginal utility. Moreover, larger value of a_c makes the utility function steeper with respect to QoS measures. In other way, the slope of the utility function is steeper with larger a_c at low value of X_{cij}. Therefore, for this particular utility function, the slope of a user's utility plays an important role in the scheduling decision of RBs. From the perspective of economics, this type of utility function has another definition. If the users with low QoS measure are given higher priority in gaining additional resource, this maximizes the social welfare as well as fairness of the system.

In order to fully explain the utility function, we need to specify the QoS measures, i.e., $\{X_{cij}^z\}_{z=1}^{m_{ci}}$. The QoS measures further require to define the traffic types of which they belong to. We consider, the network has the similar types of traffic as [130] mentioned in their work. Since the traffic types are similar, we adopt similar form of QoS definitions. However, our QoS measure is the quantity when we schedule the RBs instead of the users. Hence, we reiterate the QoS metrics of each traffic type with their exact interpretation in the following.

1. **Traffic with minimum throughput requirement:** In order to ensure minimum throughput, we need to design QoS measure so that the scheduler gives higher priority to the user with lower throughput. While allocating RB j, let denote the average throughput achieved by user (c, i) up to time t is $\bar{\zeta}_{cij}(t)$. The maximum incoming data rate for this user is ζ_{ci}^{max}, whereas the minimum required one is ζ_{ci}^{min}. As we discussed before, the property of our utility function is such that the scheduler gives priority to the user with lower QoS measure. Therefore, the QoS metric of user (c, i) for this type of traffic at time t while allocating RB j has been defined as $X_{cij}^1(t) = \left(\mu_{ci}^1 - \frac{\zeta_{ci}^{min}}{\zeta_{cij}(t)}\right)$, where $0 \leq \mu_{ci}^1 \leq 1$. Smaller value of μ_{ci}^1 gives more weight to this metric. This metric is also termed as fairness measure for this type of traffic. If any user obtains lower throughput

when the scheduler tends to allocate one RB, because of the lower value of this metric, the scheduler is forced to give provision to this user. Consequently, total utility of the system is maximized and the welfare of the system is maintained.

2. **Traffic with bounded delay constraint:** There are some traffics (e.g., audio, VoIP) which have some certain delay constraint. In order to design the metric for this class of traffic, let denote packet HOL delay of user (c, i) is $D_{cij}(t)$ at time instant t while making the scheduling decision of RB j. The maximum delay bound that user (c, i) can tolerate is D_{ci}^{max}. Similar to the case of traffic with throughput requirement, this metric represents the fairness. We define this metric as $X_{cij}^1(t) = \left(\mu_{ci}^1 - \frac{D_{cij}(t)}{D_{ci}^{max}}\right), 0 \leq \mu_{ci}^1 \leq 1$. Lower value of μ_{ci}^1 gives more weight to this metric. If any user (c, i) of this traffic type starts to experience higher packet delay, $X_{cij}^1(t)$ appears to get lower value while allocating RB j, and hence the scheduler gives more priority to that user.

3. **Best effort traffic:** Best effort traffic has usually lower priority compared to other traffic types described earlier. If any user receives considerably lower throughput comparing with other users in the system, in order to ensure fair distribution of resource, we need to design a metric for this type of traffic. Denote the measure $X_{cij}^1(t) = \left(\frac{\bar{\zeta_{cij}}(t)}{\max\limits_{(c,i)} \bar{\zeta_{cij}}(t)} - \mu_{ci}^1\right), 0 \leq \mu_{ci}^1 \leq 1$. The starvation of the users with lower average throughput results in unfairness. Hence, by serving those users at some point, the scheduler maintains the social welfare of the system. If the scheduler would serve a user with higher average throughput, it might further increase that individual user's throughput, however that contribution is lower towards the system compared to the case when the scheduler would serve a user with lower average throughput. The role of μ_{ci}^1 is the weight

for this measure, and larger value gives additional weight to this metric.

4. **Traffic with minimum throughput and bounded delay requirements:** If the traffic has both minimum throughput requirement and bounded delay constraint, the QoS measure for this type of traffic can be defined as $X^1_{cij}(t) = \left(\mu^1_{ci} - w_t \frac{\bar{\zeta}_{cij}(t)}{\zeta^{min}_{ci}} - w_d \frac{D_{cij}(t)}{D^{max}_{ci}}\right)$. Similar to other types of traffic discussed above, lower value of μ^1_{ci} gives higher priority to this QoS factor. Whether we want to give priority to delay bound or minimum throughput of this traffic type, is determined by the values of w_d and w_t. w_d and w_t are normalized by 1, i.e., $w_d + w_t = 1$. If we want to ensure equal priority to both delay and throughput of this type of traffic, we can set $w_d = w_t = 0.5$.

No matter the traffic type, we need another QoS metric which gives priority to the user with better channel condition. We define this metric as $X^2_{cij}(t) = \left(\mu^2_{ci} - \frac{R_{cij}(t)}{(\max\limits_{(c,i)} R_{cij}(t))}\right), 0 \leq \mu^2_{ci} \leq 1$. We normalize the user's instantaneous rate on RB j by the maximum possible rate achieved using this RB. The normalized rate has been subtracted from μ^2_{ci}, because we want to make sure that the user with better instantaneous rate obtains lower quantity compared to the one with worse channel condition. μ^2_{ci} works as a penalty of not serving users with good channel condition. Larger value of μ^2_{ci} gives more weight to this QoS measure. The users with better channel condition will have lower quantitative value for metric $X^2_{cij}(t)$, and hence according to the property of the utility function, the scheduler should give more provision to those users while allocating RB j at time t.

So, we have concluded that we have two QoS measures: one for ensuring fairness of specific traffic type, $X^1_{cij}(t)$ and another one is common to all users $X^2_{cij}(t)$ for ensuring the provision when the users have better channel condition.

6.3 Solution Approach and Scheduling Algorithm

In the previous section, we have seen that the marginal utility of a user is the performance metric to maximize in order to maintain the social welfare of the system. Hence, at each scheduling epoch, we want to maximize the sum marginal utility of all users across all RBs, i.e., we want to maximize $\sum_{(c,i)\in\mathbf{M}}\sum_{j\in\mathbf{N}} exp\{-a_c\left[X^1_{cij}+X^2_{cij}\right]\}$. It implies, the objective function is to maximize $\sum_{(c,i)\in\mathbf{M}}\sum_{j\in\mathbf{N}} -a_c\left[X^1_{cij}+X^2_{cij}\right]$. We denote R^{max}_j by Γ^{max}_j. Since the objective function and the constraint in Equation 6.8 have Γ^{max}_j, it is required to add an additional constraint in the problem formulation in order to support this assignment. Taking all constraints, the resultant formulation yields

$$\max_{(\mathbf{x},\mathbf{P})} \sum_{(c,i)\in\mathbf{M}}\sum_{j\in\mathbf{N}} a_c\left(\frac{R_{cij}}{\Gamma^{max}_j}-\mu^2_{ci}-X^1_{cij}\right) \tag{6.11}$$

$$\sum_{(c,i)\in\mathbf{M}} x_{cij}\le 1,\ \forall j{\in}\mathbf{N},\ \sum_{j\in\mathbf{N}} p_{cij}\le P_{ci},\ \forall(c,i){\in}\mathbf{M}$$

$$\text{if } x_{cin}=1\ \&\&\ x_{ci(n+1)}=0, x_{cij}=0, n+2{\le}j{\le}N$$

$$\text{if } x_{cin}=1\ \&\&\ x_{ci(n-1)}=0, x_{cij}=0, 1{\le}j{\le}n-2$$

$$\Gamma^{max}_j-R_{cij}\le\epsilon\Gamma^{max}_j,\ R_{cij}\le\Gamma^{max}_j.$$

The problem in Equation 6.11 has no duality gap and we can solve it by formulating it as a dual problem with the associated dual variables $\boldsymbol{\alpha}=(\alpha_{ci})_{(c,i)\in\mathbf{M}}$ for the constraint in Equation 6.2, $\boldsymbol{\beta}=(\beta_j)_{j\in\mathbf{N}}$ for the constraint in Equation 6.3 and $\boldsymbol{\gamma}=(\gamma_{cij})_{(c,i)\in\mathbf{M},j\in\mathbf{N}}$ for the constraint in Equation 6.9 and $\boldsymbol{\delta}=(\delta_{cij})_{(c,i)\in\mathbf{M},j\in\mathbf{N}}$ for the last supportive constraint. The resultant Lagrangian looks like

$$L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\mathbf{x},\mathbf{P}) := \sum_{(c,i)\in\mathbf{M},j\in\mathbf{N}} a_c\left(\frac{R_{cij}}{\Gamma_j^{max}} - \mu_{ci}^2 - X_{cij}^1\right) \tag{6.12}$$
$$+\sum_{(c,i)}\alpha_{ci}\left(P_{ci}-\sum_j p_{cij}\right)+\sum_j\beta_j\left(1-\sum_{(c,i)}x_{cij}\right)$$
$$+\sum_{(c,i)}\sum_j\gamma_{cij}\left(\epsilon\Gamma_j^{max}-\Gamma_j^{max}+R_{cij}\right)+\sum_{(c,i)}\sum_j\delta_{cij}\left(\Gamma_j^{max}-R_{cij}\right).$$

From the duality theory, the optimal solution to the problem in Equation 6.12 is given by

$$\min_{(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta})>=0}\ \max_{(\mathbf{x},\mathbf{P})} L\left(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\mathbf{x},\mathbf{P}\right). \tag{6.13}$$

In order to solve this problem, first we find the optimal value of $\mathbf{x}$ and $\mathbf{P}$ given fixed values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. Once we obtain $\mathbf{P}$, we rearrange the Lagrangian in such a way that it becomes the function of mutually exclusive per user cost function, denoted by β_{cij}. The optimal value of β_j is the maximum possible value of β_{cij} over all users for RB j taking the RB contiguity constraint in Equation 6.3 into account. Because of the RB contiguity constraint, multi-user diversity of OFDMA-based systems may not be achievable. Finally, the optimal values of $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are obtained by the help of subgradient-based numerical search method. Taking the derivative of Equation 6.12 with respect to p_{cij}, and following the K.K.T condition, we obtain optimal p_{cij}, i.e.,

$$p_{cij}^* = x_{cij}\left[\frac{a_c(1+\gamma_{cij}-\delta_{cij})}{\alpha_{ci}} - \frac{1}{e_{cij}}\right].$$

Substituting $\mathbf{p}^*$ into Equation 6.12, we obtain

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x}) := \sum_{(c,i)} \sum_{j} x_{cij} \left(f(\alpha_{ci}, \gamma_{cij}, \delta_{cij}, e_{cij}) - \beta_j \right) \tag{6.14}$$
$$+ \sum_{(c,i)} \alpha_{ci} P_{ci} + \sum_{j} \beta_j + \sum_{(c,i)} \sum_{j} \gamma_{cij} \Gamma_j^{max} (\epsilon - 1) + \sum_{(c,i)} \sum_{j} \delta_{cij} \Gamma_j^{max},$$

where $f(\alpha_{ci}, \gamma_{cij}, \delta_{cij}, e_{cij}) = a_c \left(\frac{h_{cij}}{\Gamma_j^{max}} - \mu_{ci}^2 - X_{cij}^1 \right) + \gamma_{cij} h_{cij} - \delta_{cij} h_{cij} + \frac{1}{e_{cij}} - \frac{a_c(1+\gamma_{cij}-\delta_{cij})}{\alpha_{ci}}$, $h_{cij} = \log \left(\frac{a_c e_{cij}(1+\gamma_{cij}-\delta_{cij})}{\alpha_{ci}} \right)$, $\Gamma_j^{max} = \max_{(c,i)} h_{cij}$. We denote the (c,i)th user's cost function on RB j $f(\alpha_{ci}, \gamma_{cij}, \delta_{cij}, e_{cij})$ by β_{cij}. Given that $x_{cij} \in [0,1]$, the optimal value of $\mathbf{x}$ is obtained by following the procedure below. For the sake of procedure, we copy the elements of set $\mathbf{N}$ in another set $\mathbf{N}'$.

1. First, for each RB $j \in \mathbf{N}'$, find the best RB metric among all users and denote it by $\tilde{\beta}_j = \max_{(c,i) \in \mathbf{M}} \beta_{cij}$. Second, find a RB permutation $\{\nu_j\}_{j \in \mathbf{N}}$ such that $\tilde{\beta_{\nu_1}} >= \tilde{\beta_{\nu_2}} >= \cdots >= \tilde{\beta_{\nu_N}}$. Select the RB with index ν_1 and its designated user $(c,i)|_{\nu_1}$ (to which it obtains the maximum value of the cost function), check whether the selected RB and its designated user meets the RB contiguity constraint. It means, if the designated user has already some RBs allocated, selected RB should be the contiguous to its allocated RB block; otherwise selected RB should be belonged to the designated user without any hesitation. If the RB with index ν_1 fails to satisfy the RB contiguity constraint, then the RB with index ν_2 is chosen and this operation continues until the selected RB satisfies the RB contiguity constraint. We denote finally selected RB as ν^* and its designated user is $(c,i)|_{\nu^*}$. There might be ties in this assignment which can be resolved by some inefficient search. Therefore, optimal $\mathbf{x}$ for RB ν^* is obtained as follows.

$$x^*_{cij} = \begin{cases} 1 & \text{if } j = \nu^*, (c,i) = (c,i)|_{\nu^*} \\ 0 & \text{if } j = \nu^*, \{(c,i) {\in} \mathbf{M} | (c,i) {\neq} (c,i)|_{\nu^*}\} \end{cases}.$$

2. Since the Lagrangian is a sum of the users' cost functions β_{cij}, we can minimize $L(.)$ over $\boldsymbol{\beta}$ for the given values of $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ by setting $\beta^*_{\nu^*} = \tilde{\beta_{\nu^*}}$. More than one user can obtain the value $\beta^*_{\nu^*}$, however the ties can be broken arbitrarily without losing the optimality.

3. Remove RB ν^* from set $\mathbf{N}'$.

Substituting optimal $\mathbf{x}$ and $\boldsymbol{\beta}$ into the Lagrangian, the resultant Lagrangian yields

$$L(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \sum_{(c,i)} \sum_{j} \left[\beta_{cij} - \beta^*_j\right]^+ + \sum_{(c,i)} \alpha_{ci} P_{ci} + \sum_{j} \beta^*_j + \sum_{(c,i)} \sum_{j} \gamma_{cij} \Gamma^{max}_j (\epsilon - 1) + \sum_{(c,i)} \sum_{j} \delta_{cij} \Gamma^{max}_j. \tag{6.15}$$

Notice that the Lagrangian in Equation 6.15 is the function of $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. Now, the optimal values of $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ can be obtained by minimizing $L(.)$, and this is the optimal solution of Equation 6.15. We have adopted the subgradient-based search approach in order to obtain optimal $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. The subgradient search requires the following updates in each iteration

$$\alpha_{ci}(t+1) = \alpha_{ci}(t) - \kappa(t) \left(P_{ci} - \sum_{j} p^*_{cij}(t) \right) \tag{6.16}$$

$$\gamma_{cij}(t+1) = \gamma_{cij}(t) - \kappa(t) \left(\epsilon \Gamma^{max}_j(t) - \Gamma^{max}_j(t) + h_{cij}(t) \right) \tag{6.17}$$

$$\delta_{cij}(t+1) = \delta_{cij}(t) - \kappa(t) \left(\Gamma^{max}_j(t) - h_{cij}(t) \right). \tag{6.18}$$

The step size $\kappa(t)$ in iteration $t+1$ is given by

$$\kappa(t) = \frac{\tilde{L} - L\left(\boldsymbol{\alpha}(t), \boldsymbol{\gamma}(t), \boldsymbol{\delta}(t)\right)}{\left|\frac{dL(.)}{d\boldsymbol{\alpha}(t)}\right|^2 + \left|\frac{dL(.)}{d\boldsymbol{\gamma}(t)}\right|^2 + \left|\frac{dL(.)}{d\boldsymbol{\delta}(t)}\right|^2}, \tag{6.19}$$

where $\left|\frac{dL(.)}{d\boldsymbol{\alpha}(t)}\right| = \sqrt{\sum_{(c,i)\in\mathbf{M}} \left(P_{ci} - \sum_j p^*_{cij}(t)\right)^2}$,
$\left|\frac{dL(.)}{d\boldsymbol{\gamma}(t)}\right| = \sqrt{\sum_{(c,i)\in\mathbf{M}} \sum_{j\in\mathbf{N}} \left(\epsilon\Gamma_j^{max}(t) - \Gamma_j^{max}(t) + h_{cij}(t)\right)^2}$, and
$\left|\frac{dL(.)}{d\boldsymbol{\delta}(t)}\right| = \sqrt{\sum_{(c,i)\in\mathbf{M}} \sum_{j\in\mathbf{N}} \left(\Gamma_j^{max}(t) - h_{cij}(t)\right)^2}$. $\tilde{L}$ is the estimate of the Lagrangian determined from the previous iterations. Given that $P = \max_{(c,i)\in\mathbf{M}} P_{ci}$ and $e^{max} = \max_{(c,i)\in\mathbf{M}} \max_{j\in\mathbf{N}} e_{cij}$, for this problem, each element of $\frac{dL(.)}{d\boldsymbol{\alpha}}$ is bounded within the range $[0, P]$ and that of vectors $\frac{dL(.)}{d\boldsymbol{\gamma}}$, $\frac{dL(.)}{d\boldsymbol{\delta}}$ is within $[0, (1 + loge^{max}P)]$. Under this statement, the optimal values of the Lagrangians are achievable within the finite number of iterations. The detailed procedure of achieving convergence is given in Exercise 6.3.2 of [131]. This problem has $M + 2MN$ number of dual variables, and it requires several thousands of iterations to achieve convergence. Moreover, in each iteration, there is an issue of resolving ties which requires some inefficient search. Given these disadvantages, this procedure as a scheduler may not be efficient to implement in the fast time scale. Hence, we have proposed a suboptimal algorithm presented in *Algorithm 7*.

In *Algorithm 7*, $g_{ci}(j)$ is given by

$$g_{ci}(j) = \sum_{n\in\Omega'_{ci}(j-1)} log(1 + p^*_{cin}e_{cin}) - \sum_{n\in\Omega_{ci}(j-1)} log(1 + p^*_{cin}e_{cin}), \tag{6.20}$$

where $\Omega'_{ci}(j) = \Omega_{ci}(j-1) \cup l_{ci}(j)$. $l_{ci}(j)$ is the best unallocated RB for user (c,i). p^*_{cij} is the power on RB j after doing the power control on the set $\Omega_{ci}(j)$ or $\Omega'_{ci}(j)$ for user (c,i). Power control of user (c,i) on the set $\Omega_{ci}(j)$ is equivalent to solving the following optimization problem

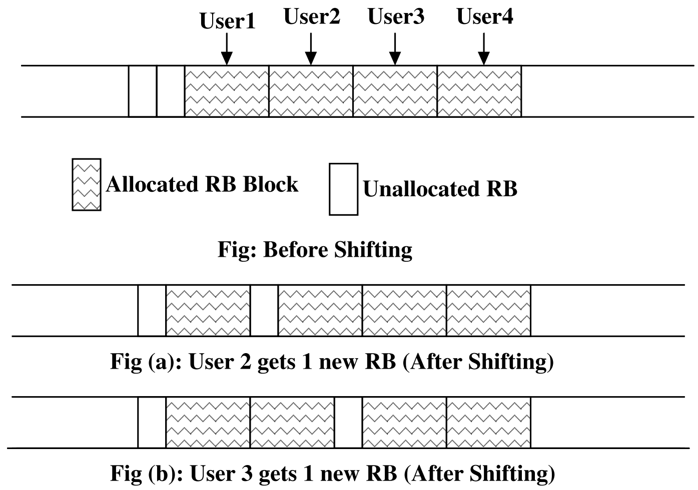


Figure 6.2: Sample subframe structure for stage 2 of Algorithm 7.

$$\underset{\sum_{n \in \Omega_{ci}(j)} p_{cin} = P_{ci}}{\arg\max} \sum_{n \in \Omega_{ci}(j)} \log\left(1 + p_{cin} e_{cin}\right). \tag{6.21}$$

The problem in Equation 6.21 can be solved by dual formulation which we have given in Section 3.5.4 of Chapter 3. The complexity of this operation is found to be $|\Omega_{ci}(j)|$. Furthermore, in the 1st stage of the algorithm, $\Gamma_{ci}(j) = g_{ci}(j)$. In the 2nd stage, we define $U_s(j)$ as $\sum_{s \in L_{sj}} \sum_{(c,i) \in L_j} -a_c \left(X_{ci}^1 + X_{ci}^2\right)$, and the parameters inside the sum term are determined presuming the network status at the beginning of current scheduling epoch.

The worst case computational complexity of step 5 is $O(MNlogN)$. Other steps, such as steps 8, 10 have the worst case complexity of $O(M)$, which is dominated by step 5. Hence, total computational complexity of the algorithm before step 14 (or stage 2) is $O(MN^2logN)$. For the 2nd stage of the algorithm, we discuss its worst case complexity as follows. In the worst case, from the 1st stage, 1 user gets 1 RB and M RBs of all users are adjacent. Therefore, the remaining number of unallocated

Algorithm 7 Suboptimal RB allocation algorithm.

ϵ is tolerable rate loss percentage and set by the service provider.
Set $j := 0$, $\Omega_{ci}(j) := \emptyset$ for each user (c,i).
while $j < N$ **do**
  Set $j := j + 1$.
  Get the best RB index $l_{ci}(j)$ for each user (c,i).
  If user (c,i) had already RBs allocated, $l_{ci}(j)$ is selected from at most 2 contiguous RBs, otherwise $l_{ci}(j)$ is arbitrarily chosen.
  Calculate metric $g_{ci}(j)$ and $\Gamma_{ci}(j)$ for each user (c,i).
  Calculate $\Gamma^{max}(j) := \max\limits_{(c,i)} \Gamma_{ci}(j)$.
  Calculate $H := \epsilon\Gamma^{max}(j)$.
  Determine users $(c,i) \in L_j$ so that $(\Gamma^{max}(j) - \Gamma_{ci}(j)) <= H$.
  Find $(c,i)^* := \arg\max\limits_{(c,i)\in L_j} a_c \left(\frac{g_{ci}(j)}{\Gamma^{max}(j)} - \mu_{ci}^2 - X_{cij}^1 \right)$.
  Assign the jth resource block to user $(c,i)^*$

$$\Omega_{ci}(j) := \begin{cases} \Omega_{ci}(j-1) \cup l_{ci}(j) & \text{if } (c,i) = (c,i)^* \\ \Omega_{ci}(j-1) & \text{Otherwise} \end{cases}.$$

end while
repeat
  Take one unallocated RB j.
  Let the users set on the right or left of RB j is L_j (presented in Figure 6.2).
  Obtain the base cumulative rate $\chi(j) = \sum_{(c,i)\in L_j} R_{ci}$.
  Obtain the base cumulative utility $U(j) = \sum_{(c,i)\in L_j} U_{ci}$.
  We index shifting operation by s and set holding all indexes is L_{sj}.
  Provide edge users in set L_j (e.g., user 1 in Figure 6.2) available RBs if necessary, $\forall s \in L_{sj}$.
  Determine the cumulative rate $\chi_s(j)$ and the cumulative utility $U_s(j)$, $\forall s \in L_{sj}$.
  Determine shifting set L'_{sj} so that $(\chi^{max}(j) - \chi_s(j)) <= H$ and $U_s(j) >= U(j)$.
  Find optimal s, $s^* := \arg\max_{s \in L'_{sj}} U_s(j)$.
until No Improvement is possible

RBs is $N - M$, and the loop of the 2nd stage runs $N - M$ times. Inside the loop, we may need to shift M times, and each shifting requires power control and other primitive steps, which are of $O(1)$ complexity. Hence, the worst case complexity of the 2nd stage is $(N - M)M$, which is again dominated by the complexity of the 1st stage.

6.4 Characteristics of Proposed Scheduling Scheme

In this section, we first prove that our scheduling scheme achieves much fairer resource distribution compared to [130]. Second, we discuss the impact of opportunity cost function on the scheduling outcome, and show that for certain value of H, the proposed scheduling scheme can bridge between the MT (which has the best global performance) and PF (which is well-known for proportional fairness) schemes. In the last paragraph, we have briefly discussed the practicality of our scheduling scheme with respect to other schedulers implemented in practice.

Conjecture 6.1: Our scheduling scheme (both the optimal and *Algorithm 7*) is more efficient in terms of achieving fairness compared to [130].

Proof: Consider an epoch where there is 3 RBs and 3 remaining users available to schedule. User 1 has the best SNR condition for all RBs, i.e., $e_{1j} >> e_{2j}, e_{1j} >> e_{3j}, \forall j \in [1, 3]$. User 2 has almost similar SNR condition compared to user 3, however slightly better. The packet HOL delay for user 1 is much smaller compared to user 2, i.e., $D_1 << D_2$ (or $\zeta_1 >> \zeta_2$) and $D_2 \approx D_3$ (or $\zeta_2 \approx \zeta_3$). If we would apply the scheduling technique presented in [130], it is likely that the rate obtained by user 1 is much higher compared to that of user 2 or 3 due to the favorable physical layer condition. In our simulation, we observe, if any user has favorable channel condition, it is more likely, that user obtains more RBs comparing with others. Hence, in our scenario, there is a chance that user 1 gets 2 RBs and user 2 gets 1. However, the result is not fair given the packet HOL delay or throughput. With our scheme, possible RB allocations are: **1)** user 1 may get first RB due to its good SNR condition on all RBs, then RB 2 is assigned to user 2 due to its better SNR condition on the remaining RBs and almost similar packet HOL delay or throughput compared to user 3, finally user 3 gets the remaining last unallocated RB. **2)** user 1 may obtain the 1st

RB due to the similar reason described above, user 2 gets remaining 2 RBs because of its SNR condition and the worst QoS performance compared to user 3.These 2 scheduling decisions are considered as fair compared to that by [130].

In order to show that there is no scenario that the scheduling technique [130] outperforms our scheme, we introduce a counter example of this example. Similar to this example, consider user 1 has the best channel condition on all RBs compared to the other two users. Unlike before, the average packet HOL delay or throughput of this user is much worse compared to the other two users. The channel condition of user 2 is little better compared to user 3 on all RBs, however much worse compared to user 1. Furthermore, the packet HOL delay or throughput for these two users are almost similar, however better than user 1. For the similar reason explained in the previous example, by the scheduler [130], user 1 obtains first 2 RBs and user 2 obtains the 3rd RB. On the other hand, by our scheduling scheme, two possible RB allocations are: **1)** user 1 gets first 2 RBs because of its best channel condition and worse packet HOL delay or throughput. The rest 1 RB will be allocated to user 2 because of its slightly better channel condition compared to user 3. **2)** user 1 obtains all 3 RBs. The scheduling decisions by both [130] and our scheduling scheme are considered as fair. Hence, there is no scenario for which the scheduler [130] can perform better than our scheduler. Our scheduling scheme always achieves better or as good as performance in terms of fairness comparing with [130].

Lemma 6.1: Under the special case ($H = 0$), *Algorithm 7* converges to the MT scheme.

Proof: In the 1st stage of *Algorithm 1*, the MT scheme assigns RB $l_{ci}(j)$ to user (c, i) when it achieves the maximum quantity for $g_{ci}(j)$ compared to other users; whereas for the 2nd stage, it selects shifting operation which has utility greater than the base cumulative utility and have the maximum value comparing with other shifting operations. Set $H = 0$, at the 1st stage, our scheduler selects only one user for set L_j whose $g_{ci}(j)$ has the maximum quantity in order to satisfy the condition $\Gamma^{max}(j) - \Gamma_{ci}(j) <= 0$. Afterward, since L_j has only one user, eventually this user will be picked for RB $l_{ci}(j)$. For the 2nd stage, similar situation happens, L'_{sj} contains only the shifting operation which incurs the highest utility and has larger value compared to the base cumulative utility. Hence, the converging trend of our algorithm towards the MT scheme has been proved.

Lemma 6.2: For the special value (the maximum achievable rate, $Rate_{Max}$) of H, *Algorithm 7* is transformed to the PF scheme.

Proof: At scheduling epoch t, while scheduling RB j, the PF algorithm assigns RB $l_{ci}(j)$ to user (c,i) if its metric $\log_2(1+\frac{g_{ci}(j)}{\bar{\zeta}_{cij}})$ obtains larger quantity compared to other users. Note that the PF metric is an increasing function of $g_{ci}(j)$ and a decreasing function of its long term throughput $\bar{\zeta}_{cij}$. Consider about our scheme with $H = R_j^{max}$, set L_j contains all potential users in the system. And, then, the scheduler picks the user whose marginal utility, i.e., slope obtains the highest quantity. The slope of a user is the increasing function of $g_{ci}(j)$ and has a decreasing trend with respect to its long term throughput $\bar{\zeta}_{cij}$. Mathematically, we can equate both functions in order to find the instantaneous value for the constants. Hence,

$$\log_2(1+\frac{g_{cij}}{\bar{\zeta}_{cij}}) \;=\; -\,a_c\left(\mu_{ci}^1 - \frac{\zeta_{ci}^{min}}{\bar{\zeta}_{cij}} + \mu_{ci}^2 - \frac{g_{cij}}{\Gamma_j^{max}}\right).$$

Simplifying this, we obtain

$$\mu_{ci}^2 + \mu_{ci}^1 \;=\; -\frac{1}{a_c}\log_2(1+\frac{g_{cij}}{\bar{\zeta}_{cij}}) + \frac{\zeta_{ci}^{min}}{\bar{\zeta}_{cij}} + \frac{g_{cij}}{\Gamma_j^{max}}.$$

Therefore, we can conclude, at scheduling instant t while allocating RB j, if the value of $\mu_{ci}^2 + \mu_{ci}^1$ is equivalent to the right hand side of above equation (for user (c, i)), and if we ignore the opportunity cost function, over the infinite time, our scheduling algorithm converges to the PF scheme. In the similar manner, for the 2nd stage of *Algorithm 7*, we can prove that at scheduling epoch t, certain value of constant $\sum_{(c,i)\in L_j} \mu_{ci}^2 + \mu_{ci}^1$ results in the asymptotic convergence towards the PF scheme.

Proofs 6.1 and 6.2 remind us that the presented scheduling algorithm in this thesis is generalized. In general, this scheduler works in between these two extreme cases which has been proved in the next section. Since the optimal solution is an iterative procedure, for this, there is no closed form proof for *Lemma 6.1* and *Lemma 6.2*. However, for the same condition of these two Lemmas, the optimal solution converges to the MT and PF schemes, respectively. Adjusting the parameter of the opportunity cost function, we can convert this scheduler to two conventional extreme cases of scheduling scheme.

Furthermore, we have noticed that Huawei [132] has implemented an enhanced PF scheduler for the GBR traffic. Packet delay budget of different GBR traffic and aggregate RB quality are considered while designing this scheduler. For the uplink scheduling, this scheduler first calculates priority metric of each user which is a function of average packet delay and approximate average channel quality over all RBs of that user, and then based on the priority, it allocates RBs among those users. The priority metric is calculated using MW formula, and M-LWDF is similar version of the MW rule. Whereas, our scheduler does one by one RB allocation based on the instantaneous RB's channel quality and average packet delay, which is apparently more dynamic. Moreover, our scheduler can deal the traffic with throughput requirement, and the design of the scheduler is flexible for diverse QoS (e.g., packet loss ratio, packet jitter) oriented traffic. Besides, the scheduler designed by Ericsson [133] applies conventional traffic policing and shaping concept while allocating RBs among different QoS-based traffic. Policing ensures that the users do not get the agreed upon configured rate, whereas traffic shaping makes sure that the users get at least minimum settled QoS specified in their agreement with the vendor. Based on the traffic policing and shaping mechanisms, our scheduler is dynamic and can achieve the similar level of performance as the scheduler equipped with these policies.

6.5 Performance Evaluation

In this section, we will evaluate *Algorithm 7* for four classes of traffic. First, we outline the simulation methodology we have adopted. Then, the simulation results are split in two parts for Homogeneous and Heterogeneous traffic, respectively.

6.5.1 Simulation Methodology

For setting up the network, we put the base station at the center of a cell, user nodes at different distance surrounding the base station. Cell radius is assumed as 1 km. We run the simulation over 5000k TTIs. One TTI is equivalent to 1ms and it consists of 25 RBs. Each RB is analogous to 12 subcarriers [6]. These total 25×12 subcarriers are spread over 5 MHZ bandwidth. The theoretical limit [82] of the channel capacity is given by $\beta = \frac{-1.5}{\ln(5P_b)}$, where P_b denotes the BER. BER for the channel is configured as 10^{-6}. Each user's maximum power is set as 220 mW. In order to calculate the path loss effect of the channel, we assume the reference distance as 1 km and the SNR for this reference distance is 28 dB. The reference shadowing effect is the log normal distribution with variance 3.76. With this variance, log-normal shadowing power is determined as 10.6 dB according to [134]. Rayleigh fading effect is captured by a parameter a such that $E[a^2] = 1$. The channel gain for a particular user (c, i)

over RB j is computed by

$$G_{cij,dB} = (-\kappa - \lambda \log_{10} d_{ci}) - \xi_{cij} + 10 \log_{10} F_{cij}. \tag{6.22}$$

In Equation 6.22, the first term κ captures propagation loss, the value of which is 128.1 dB. d_{ci} is the distance in km from user (c, i) to the base station, and λ is the path loss exponent which is set to a value 3.76. The second term ξ_{cij} captures the path loss effect for the reference distance 1 km. Whereas, the last term F_{cij} corresponds to the rayleigh fading effect. Feedback duration due to the exchange of CSI, scheduling decisions between the users and base station is considered as negligible. Perfect CSI estimation is assumed at the base station.

In order to demonstrate the ability of our uplink scheduling scheme, we need to justify that users with different QoS measures obtain expected service which has already been specified theoretically. In the previous section, we have designed the utility function with 3 different QoS measures. Now, we want to define 4 different classes for the users where each class is accompanied with one QoS measure. Four different classes are: VoIP (class 1), audio streaming (class 2), video streaming (class 3) and FTP (class 4). Class 1 has the highest priority and class 4 is the type of lowest priority. Class 1 and 2 traffics have delay bound constraint. Video streaming has minimum throughput constraint[14], whereas class 4 is the type of best effort traffic. In order to distinguish priority of different classes, we set the parameters for a_c and μ_{ci}, which are shown in Table 6.5.1.

For VoIP traffic, we have taken adaptive multi rate (AMR) codec [135] method. According to this model, packets are generated using a negative exponentially dis-

[14]There are two types of video traffic, e.g., interactive video with stringent real-time requirements; video streaming and video download with some minimum bandwidth requirement. In our simulation, each video flow is of second type.

tributed ON-OFF pattern to replicate the talk and silent duration of a VoIP call. The mean duration of ON and OFF periods are 3 s. During the ON period, in every 20 ms interval, a voice packet of 244 bits is generated. Including the compressed IP/UDP/RTP header, the data rate for each VoIP flow becomes 13.6 Kbps. According to [136], the maximum acceptable delay for voice is 250 ms. Considering the delay induced by the core network and the delay for RLC and MAC buffering, the tolerable delay at the radio interface should be at most 100 ms [6], which represents a very strict requirement. For modeling audio streaming traffic, we have also used AMR codec. The size of the packets generated for each audio streaming user is uniformly distributed between 244 and 488 bits, and hence the data rate varies from 12 Kbps to 64 Kbps. The maximum delay threshold has been set 150 ms for each user. Video streaming is modeled with a minimum data rate of 64 Kbps and a maximum data rate of 384 Kbps. The size of the packets in each video flow is uniformly distributed between 1200 and 2400 bits. Hence, the resultant each video flow generates the number of packets which is uniformly distributed between 1 and 3 at the interval of 19 ms.[15] The data rate of each user with FTP traffic is assumed as maximum 128 Kbps with the size of a packet 1200 bits. Packet generation interval of each FTP flow is 10 ms. In the simulation, we assume each user is equipped with one class of traffic and it keeps holding that flow until the simulation is finished. All the results described in the following subsection are average of 20 simulation runs. As the simulation tool, we have modified the matlab-based LTE simulator [137] with all functionalities of our scheduling scheme.

[15] Each video flow is designed following the trace file "sony" taken from "http://trace.eas.asu.edu/h264/". The statistics of this trace is: **a.** Inter-arrival time between two bursts is constant (33 ms), **b.** Burstiness of the video is determined by the size of the burst. The maximum frame burst of this video is 326,905 bytes and the minimum one is 20,209 bytes. Burstiness of our each video flow is determined by the number of packets and packet size. According to our statistics, maximum frame size of each video flow is 7200 bits and the minimum one is 1200 bits.

Table 6.1: Simulation parameters.

Traffic Type	a_c	μ^1_{ci}	μ^2_{ci}
VoIP	4	0.4	0.3
Audio	3.5	0.4	0.3
Video	3.0	0.4	0.3
FTP	2.5	0.4	0.6

6.5.2 Simulation Results

First we show the simulation results for homogeneous traffic, i.e., VoIP and traffic with throughput requirement (video streaming). With regard to this experimentation, we have compared the results obtained by our scheduling algorithm with the existing work. For VoIP traffic, we have compared our results with M-LWDF [63], EXP [54] rules and [62]. EXP and M-LWDF rules are specifically developed for the downlink scheduling of delay sensitive multimedia traffic in OFDM-based systems, whereas [62] is for the uplink scheduling scheme. Since these schemes have limitations and are not directly comparable to our scheme, we have extracted scheduling rules from those works and substitute in our algorithm in order to have valid comparison. The scheduler under M-LWDF rule assigns RB $l_{ci}(j)$ to the user $(c,i)^*$ abiding by the formula

$$(c,i)^* = \arg\max_{(c,i)\in\mathbf{N}} \frac{g_{ci}(j)}{\bar{\zeta}_{ci}} D_{cij}.$$

The scheduler with EXP rule obeys the following rule for RB $l_{ci}(j)$

$$(c,i)^* = \arg\max_{(c,i)\in\mathbf{N}} b_{ci}\exp\left(\frac{a_{ci}D_{cij}}{1+\sqrt{(1/N)\sum_{(c,i)} D_{cij}}}\right) g_{ci}(j),$$

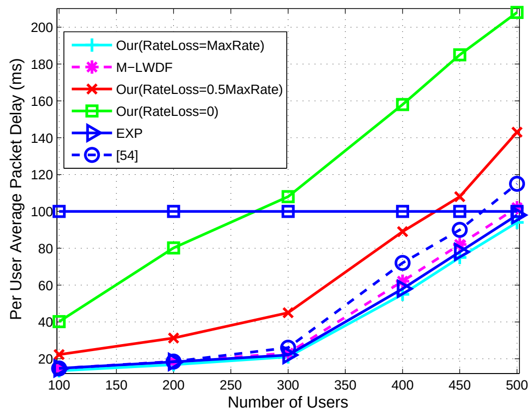
where $b_{ci} = 1/E[g_{ci}]$ and $a_{ci} = \frac{6}{D^{max}_{ci}}$.

For ensuring guaranteed bitrate, we did not find much work in the literature except [55]. Therefore, for the homogeneous video streaming setup, we have done performance comparison with [55] as well as with the MT and PF schemes.

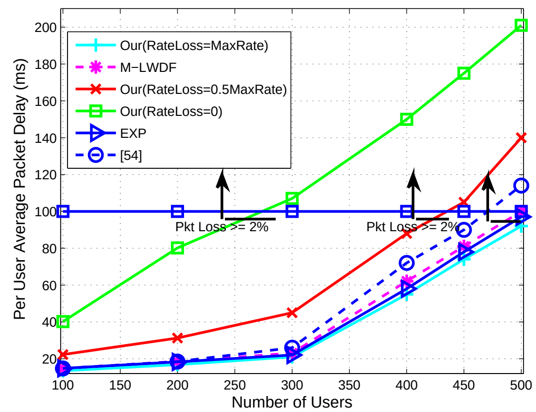
Although [54] is the solution for mixed traffic, they did not consider the traffic with guaranteed bit rate while evaluating the performance. The most recent work dealing all types of traffic is [59], and we have taken it as a performance benchmark while presenting the results of mixed traffic obtained by our scheme.

For the performance metrics, we consider average packet delay, average normalized throughput, proportion of per RB rate loss on behalf of the service provider, per RB normalized utilization, total system's effective rate, fairness in the system. For measuring fairness over all users in the system, we have used well-known fairness indicator named as Jain Fairness Index (JFI) [83]. Proportion of per RB rate loss is the indication of how much proportion of rate is sacrificed from the maximum one (that could be achieved) while taking the scheduling decision of each RB. On the other hand, per RB utilization is the measure of RB utilization over the entire simulation interval. While computing RB utilization, 0 is counted when any RB stays vacant and contributes to the average per RB utilization. Furthermore, we consider, the users are spread uniformly between 0.5 km to 0.8 km distance from the base station.

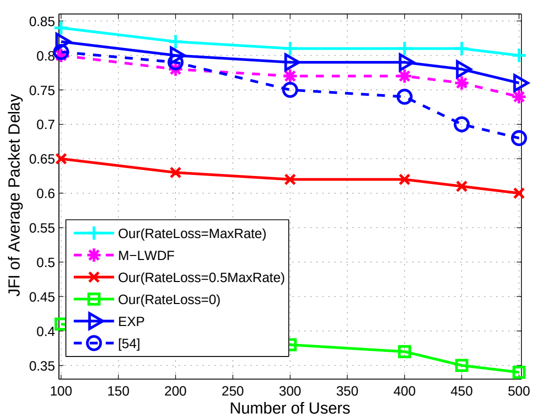
In the second part of the simulation results, we have studied the performance of the multiplexed traffic. The simulation scenario is designed in such a way that we see the strength of our scheduling scheme which can prioritize different classes of traffic and can reach to an elegant scheduling decision.



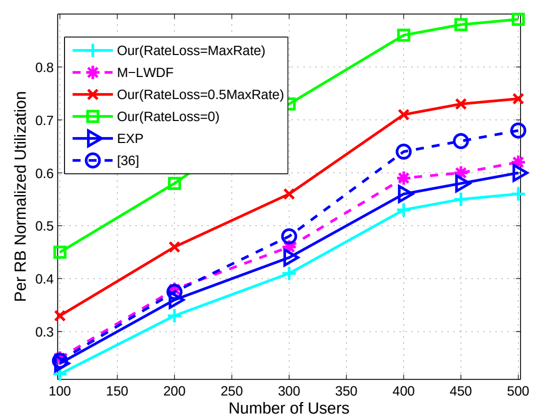
(a) Per user average packet delay comparison between our scheme and others (infinite buffer).



(b) Per user average packet delay comparison between our scheme and others (finite buffer).



(c) JFI of average packet delay comparison between our scheme and others.



(d) Per RB normalized utilization comparison between our scheme and others.

Figure 6.3: VoIP.

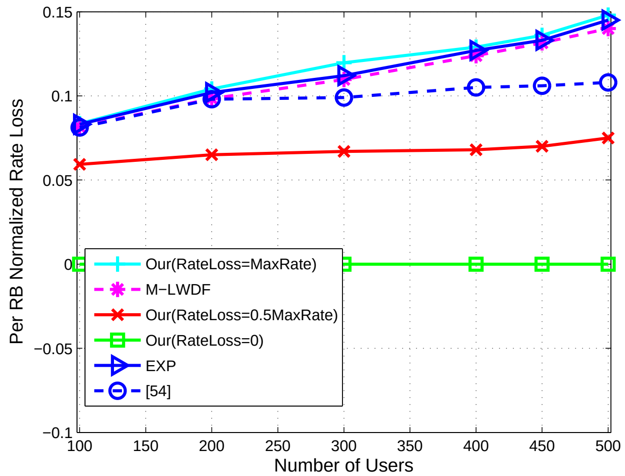


Figure 6.4: Per RB normalized rate loss comparison between our scheme and other (VoIP).

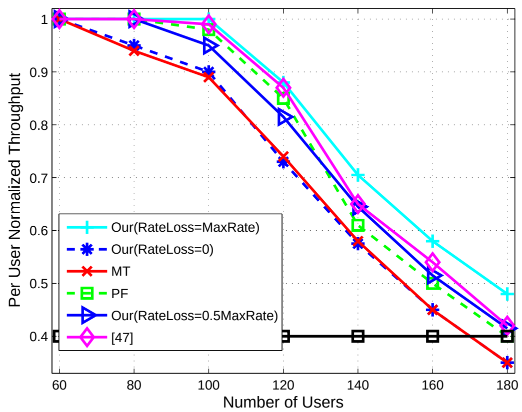
6.5.2.1 Homogeneous Traffic (*VoIP*)

Figure 6.3(a) depicts the average packet delay of VoIP traffic with respect to the total number of users in the system. First observation from this figure is, with the increased number of users, the average packet delay is increased for all cases. We have shown the results of our algorithm for different maximum tolerable rate loss, i.e., $Rate_{Max}$, $0.5Rate_{Max}$ and 0. As we mentioned before that if maximum tolerable rate loss is $Rate_{Max}$, it treats all users as if the network operator does not have problem with any rate loss, this setting treats all users equally. The figure indicates, for the parameters given in Table 6.5.1, our scheme (with the maximum tolerable rate loss $Rate_{Max}$) has better performance compared to the scheduler with EXP and M-LWDF rules. The M-LWDF and EXP rules are specifically designed for the delay sensitive multimedia network. The formula of the M-LWDF scheduler is based on the product of the packet HOL delay and PF factor. Hence, the resultant scheduling policy ensures fairness among the users. The scheduler with EXP rule is more robust than the scheduler with M-LWDF rule. This is because the exponential function grows much faster with its argument. Furthermore, the EXP scheduler also takes the overall network status into account, because the delay of considered user is somehow normalized over the sum of experienced delay of all users. Better performance of our scheme with respect to M-LWDF and EXP rules justifies its efficacy for the usage of delay sensitive VoIP traffic.[16] At low load, [62] performs almost in the similar manner compared to our algorithm with negligible deviation. However, at high load, the users with moderately better channel condition have more priority, while the users with worst channel condition almost starve, because the scheduling rule depends on the time difference between the current and recent burst. Therefore, within a fixed time period, the users with good channel condition get scheduled while keeping the users with worst channel starved. While serving these users, another burst of traffic arrives for all users and hence the delay based metric is replaced by the same value for all users including the users with worst channel. Therefore, in the second round, for the same value of delay based metric, the same set of users with good channel condition get served which results in starvation for the users with worst channel. If the delay based metric of [62] would consider the time of first burst instead of recent, it would perform better at high load. When the maximum tolerable rate loss is 0, our scheme performs poorly. This is because the scheduler cannot tolerate any rate loss and it only serves the high rate generating users which causes higher packet delay for other users. Due to the higher packet delay of the low rate generating users, the resultant average packet delay of the system gets larger. The performance of our scheduler with the maximum tolerable rate loss $0.5Rate_{Max}$ has in-between performance of other two for the similar reason. Instead of using infinite buffer, if we use finite buffer at each

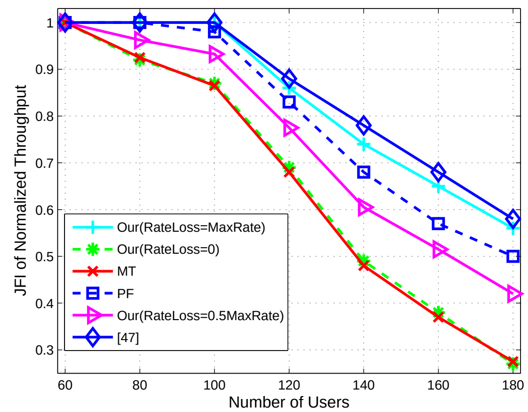
[16] By adjusting the parameters μ_{ci}^1, μ_{ci}^2 in our scheduler and the parameters a_{ci} and b_{ci} in the EXP scheduler, better or comparable performance (with respect to our scheduler) can be achievable by the scheduler with EXP or M-LWDF rule.

terminal with limited size, we observe packet loss. We define the outage of a system with VoIP users when its packet loss exceeds 2%. This is because if a user suffers at least 2% packet loss, it is likely that the packets of that user cannot be decoded. Therefore, due to the limited buffer, even though the average packet delay of the system is below the noted limit, QoS is not met for the users in the system because of more than 2% packet loss. Hence, we see in Figure 6.3(b) that the coverage of the system with finite buffer is at earlier point compared to that with infinite buffer. This justification applies for all schedulers.

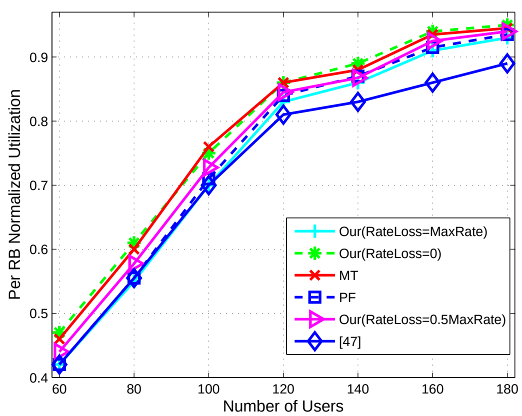
Our scheduler with the maximum tolerable rate loss $Rate_{Max}$ experiences lower average delay. It implies, the scheduler gives more priority to the users with worse channel condition compared to the scheduler with M-LWDF or EXP rule. Therefore, in terms of fairness, our scheduler with the maximum tolerable rate loss $Rate_{Max}$ outperforms the rest others as depicted in Figure 6.3(c). The utility function combined with its parameters of our scheduler is specifically designed to preserve the fairness across the system while exploiting users' varying channel quality condition. With the decrementing maximum tolerable rate loss, we see the decrementing JFI. This behavior is expected, because the scheduler picks selectively high rate generating users, and hence the scheme with the maximum tolerable rate loss 0 has the worst fairness. The metrics, such as percentage of RB utilization and rate loss are the best for this case and have been illustrated in Figures 6.3(d) and 6.4, respectively. With the decrementing tolerable rate loss, the scheduler gradually ignores the users with better channel condition and tries to serve the users whose average delay tends to deviate from the prescribed bound, and hence we see the decrementing RB utilization and incrementing rate loss.



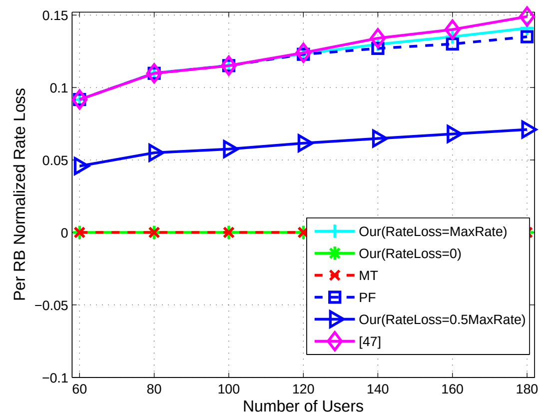
(a) Per user normalized throughput comparison between our scheme and others.



(b) JFI of normalized throughput comparison between our scheme and others.



(c) JFI of average packet delay comparison between our scheme and others.



(d) Per RB normalized rate loss comparison between our scheme and others.

Figure 6.5: Video streaming.

6.5.2.2 Homogeneous Traffic (*Video Streaming*)

In order to present results for this case, a simulation setup similar to that used in the previous subsection has been undertaken. Figure 6.5(a) shows the average normalized throughput with the increasing total number of users. It is expected that increased number of users deteriorates per user normalized throughput due to the limited number of RBs available at each scheduling epoch. In the figure, we have shown the performance of our scheme with three maximum tolerable rate loss,

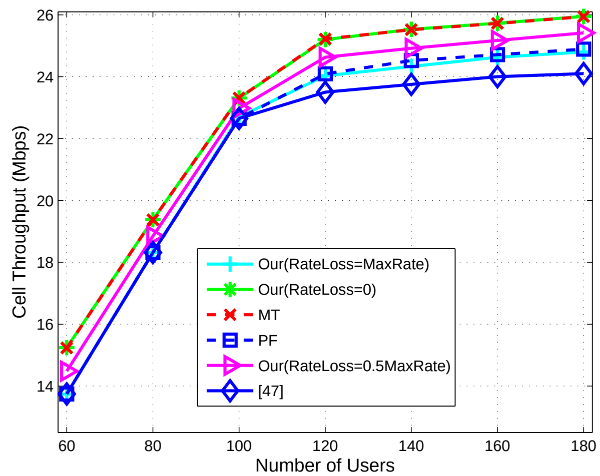


Figure 6.6: Cell throughput comparison between our scheme and others (Video).

i.e., $Rate_{Max}$, $0.5Rate_{Max}$ and 0. Our scheme with the maximum tolerable rate loss $Rate_{Max}$ has the best performance. Because of the fairness measure in terms of per user average throughput inside the exponential utility function, our scheme ensures fairness across the users. In [55], the scheduler works in both time and frequency domains. Time domain task partitions the users based on their throughput being less and greater than the noted limit. The former list has absolutely higher priority than the latter one. Moreover, the priority metric for the first list is determined from blind equal throughput, whereas the metric for the second one is the ratio of instantaneous wide band channel quality and the past average throughput. From the sorted priority list, N number of users are passed to the frequency domain scheduler which applies PF scheme in order to finalize the scheduled users. The PF scheme uses the product of the instantaneous RB quality and inverse throughput in order to ensure fairness. This scheduler pre-processes the users while giving priority to the ones with lower throughput before applying the PF technique on them. Therefore, at low load, the PF technique and [55] perform almost similarly as the higher priority list selected by the time domain scheduler is empty, and hence there is no basic performance difference between them. However, at high load, the higher priority list [55] gets bigger and bigger, and consequently it performs better compared to the PF one. It is unusual to see that the MT scheme has lower per user normalized throughput, however, there is a intuitive reason behind it. This scheduler gives higher priority to the users with better channel condition even though those contribute less towards the overall system throughput, and hence several users with worse channel remain under-provisioned. Same thing happens to our scheduler with the maximum tolerable rate loss 0. The higher rate generating users are given priority in the scheduling decision at the expense of the lower rate generating users, and hence the resultant normalized average throughput of the system is lower. If we use limited buffer at each user terminal, we notice even lower coverage for all cases because of the buffer overflow resulting in packet loss.

Unlike the results discussed in the previous paragraph, we observe better performance for the MT scheduler or our scheme with the maximum allowable rate loss 0 in terms of overall cell throughput or percentage of RB utilization. The results with this respect are given in Figures 6.6 and 6.5(c), respectively. The MT or our scheme with the maximum tolerable rate loss 0 selects only the users with better channel condition or the higher rate generating users. In either case, such behavior of the scheduler increases cell throughput or enhances per RB utilization, however at the expense of the throughput for the users with worse channel condition or the low rate generating users. Since [55] keeps busy in serving the low rate generating users at high load, per RB utilization is the worst for this scheduler. Because we give priority to the best users at each scheduling epoch, the percentage of rate loss is 0 for the MT scheme or our scheme with the maximum tolerable rate loss 0 as depicted in Figure 6.5(d). Whereas, the scheduling decision of other scheme [55] not necessarily depends on user channel condition or rate, they check per user's instant throughput and relevant constraint. And, hence, the scheduler of this policy suffers higher rate loss. The fairness in terms of JFI comparison among all schemes is given in Figure 6.5(b). Because of the exponential nature of the utility function, if any user goes under the minimum throughput, our scheduler (with the maximum tolerable rate loss $Rate_{Max}$) is forced to serve that user, and hence ensures fairness across the users while exploiting their instantaneous channel conditions. Because of the nature of the utility function, the PF scheme ensures fairness across the system, however is not as good as our scheme. At low load, the scheduler [55] behaves like the PF scheme and so thus the fairness. However, at higher load, its time domain scheduler is almost ignorant of the spectral efficiency, gives priority to the users with lower throughput, and hence achieves the highest fairness compared to all. Our scheme does not deviate much from [55] in terms of fairness. With the decreasing maximum tolerable rate loss, our scheme also suffers from fairness measure because of giving privilege to the higher rate generating users.

6.5.2.3 Heterogeneous Traffic

In order to prove that our approach can handle multiplexed traffic efficiently, we have deployed N users splitted equally and uniformly into 4 classes. Moreover, maximum tolerable rate loss is assumed as $Rate_{Max}$ and buffer size in each terminal is considered as infinite. Figures 6.7 and 6.8 compare the average packet delay of VoIP and audio streaming users, respectively with that of [59]. In the same figure, we have shown the results from the homogeneous setup obtained by our scheme. As VoIP has higher priority than the audio streaming, in the heterogeneous setup, the VoIP users will always incur lower packet delay comparing with the audio streaming users. Moreover,

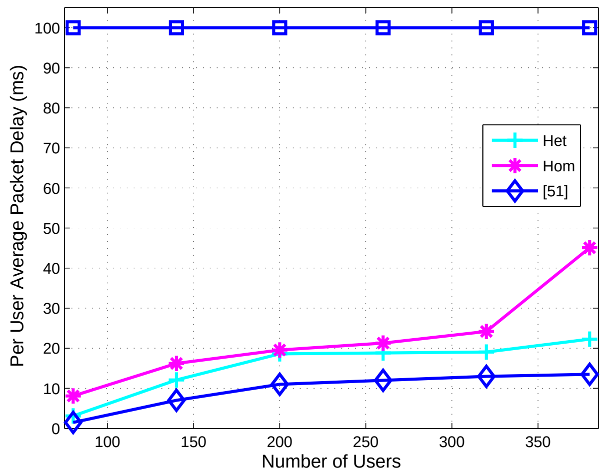


Figure 6.7: Per user average packet delay comparison between homogeneous and heterogeneous setup (VoIP).

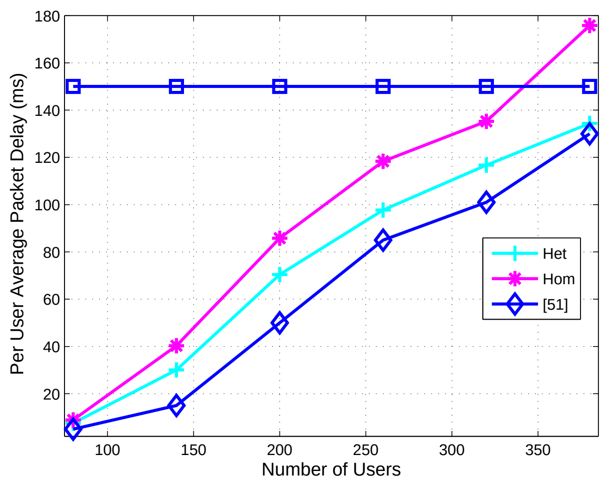


Figure 6.8: Per user average packet delay comparison between homogeneous and heterogeneous setup (Audio).

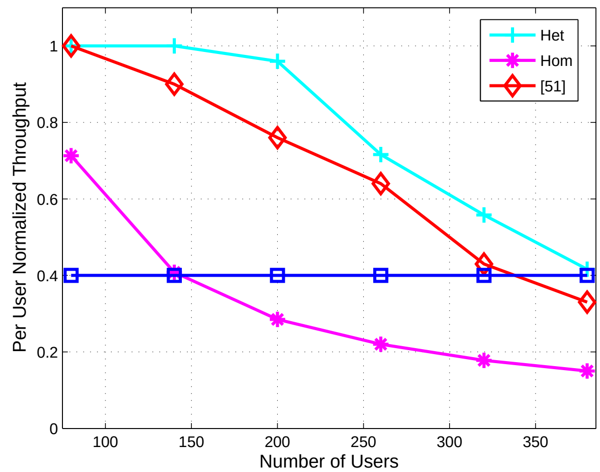


Figure 6.9: Per user normalized throughput comparison between homogeneous and heterogeneous setup (Video).

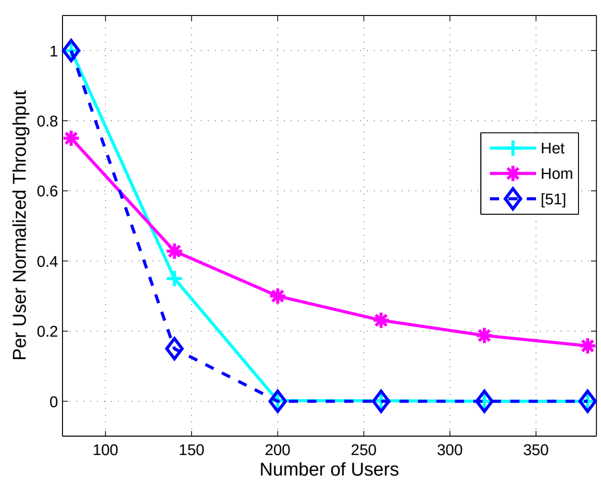


Figure 6.10: Per user normalized throughput comparison between homogeneous and heterogeneous setup (FTP).

the heterogeneous VoIP users always have lower average packet delay than the case with the homogeneous setup. This is because total users in the multiplexed case is equally divided among the rest 3 lower priority classes, and hence there are fewer VoIP users in this case comparing with the single type one. [59] always gives the highest priority to the VoIP users whenever they have data in the queue no matter the accumulated delay of the users is much lower than the bounded limit. On the other hand, the audio streaming users under multiplexed setup not necessarily always have lower packet delay comparing with that of homogeneous case although there are some lower priority users along with them. Since the data rate of each audio streaming user is larger than that of a VoIP user, at low load, when the network is unsaturated, multiplexed less number of audio streaming users have enough resource to get their packets transmitted. However, as we increase the load, the number of VoIP users increases, they may occupy the entire resources of the network, in that case, we will be able to see higher packet delay for the multiplexed case comparing with the single class case. Audio streaming is the second highest priority class and the data rate of the users with this class is much larger than that with VoIP, the scheduler in [59] apparently serves the users of this type whenever they have packets in the queue no matter their packet HOL delay is lower or higher than the noted limit. This reason results in lower average packet delay compared to our scheme. However, when the network will be saturated with the VoIP users, the audio streaming users will achieve closer or worse performance compared to our algorithm.

Figures 6.9 and 6.10 depict the average normalized throughput for the video streaming and FTP users, respectively. Similar to the previous figures, here, we have shown the results for the homogeneous setup achieved by our scheduler. Since video streaming is higher priority type, the users of this type incur higher throughput than

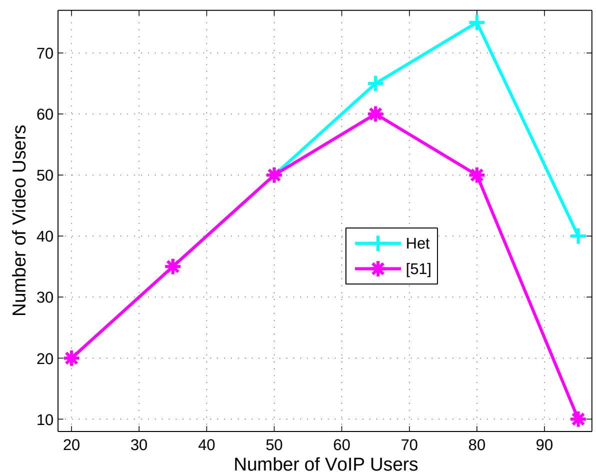


Figure 6.11: Admission region with QoS guarantee comparison between our scheme and other.

that of FTP. Similar to the audio streaming users, at low load, under multiplexed case, the video streaming users achieve better performance comparing with the homogeneous users. In addition, each video streaming user has much higher data rate comparing with the combined VoIP and audio streaming users. If we increase the number of users in the network and it goes close to the saturation with the VoIP and audio streaming users, the video streaming users will start to starve. And, at that point, we will see the poor performance of the heterogeneous video streaming users. Same reasoning applies to the FTP users, at low load, we will see better performance for the heterogeneous case. However, when all resource of the network is consumed by all higher priority traffic, the performance of the heterogeneous FTP users starts to deteriorate. Because of the blind provision towards the VoIP and audio streaming users given by the scheduler [59], the video streaming users perform poorly compared to our algorithm. For the similar reason, the throughput of FTP users is even worse comparing with our scheme.

Figure 6.11 compares the admission region of VoIP and video streaming users in the system under multiplexed setup with [59]. Since [59] redundantly gives more provision to the VoIP users, the average per user packet delay designed by them is lower than that by our scheme as depicted in Figure 6.7. It starves the video streaming users although the delay of VoIP users is lower than the required limit. Therefore, from the figure, we observe, although the number of VoIP users with QoS assurance is equal to our scheme, the number of video streaming users with guaranteed throughput is lower at increased load.

6.6 Summary of the Chapter

LTE and LTE-Advanced are specifically envisioned for the applications with high data rate. According to this standard, the spectrum is divided into time-frequency-based resource units called RBs. Varying emergent applications prompt the development of heterogeneous traffic networks with diverse QoS requirements. Besides the demands of end users, resource utilization is an important matter to consider in order to assist the network operators. In this chapter, we have presented an uplink scheduling technique for LTE heterogeneous traffic networks which is able to maintain varying QoS provision of end users while keeping the resource utilization (in RB level) of service providers in the prescribed range. In order to solve this problem, we have first formulated the problem considering all LTE standard specific constraints while capturing QoS factors of the users in a smart utility function. In addition to these, the formulation consists of a constraint which allows the service provider to keep granular RB utilization level in some certain range. Having considered the discrete nature of RB allocation, we have shown the optimal solution structure of the problem which has high computational complexity. From the guiding principles of the optimal solution structure, we have proposed a suboptimal algorithm with polynomial time complexity. Furthermore, we have evaluated our scheduler in a network which has three different types of users, i.e., traffic with delay constraint, traffic with minimum data rate requirement and best effort traffic. The results obtained from extensive simulation show that the proposed scheme exhibits the tradeoff between the fairness of end users and the resource utilization of service providers. By showing the simulation results for both homogeneous and heterogeneous traffic networks in terms of several performance metrics, we have justified the efficacy and effectiveness of our scheduling scheme envisioned to be deployed in future LTE systems.

Chapter 7

Conclusion and Future Work

In Section 7.1 of this final chapter, we briefly review our results and highlight the contributions of this thesis. In Section 7.2, we provide a few possible future works resultant from the work of this thesis.

7.1 Summary of the Thesis

This thesis is focused on the uplink scheduling and resource allocation of LTE-Advanced systems that incorporate relay(s) or carry heterogeneous traffic. The main results of each chapter are summarized as follows.

In Chapter 3, we have proposed a uplink scheduling and resource allocation scheme for DF relay (with FD functionality) equipped LTE systems. Although there are many sophisticated advancement on the uplink scheduling in relay aided systems, basic scheduling scheme is still missing. We fill that gap in the literature. Besides providing the solution for basic scheduling and resource allocation, the proposed solutions are adaptive. If the relays are in the bad locations or cannot contribute much towards maximizing the network capacity, they are recommended to be disabled. However, due to the implementation complication, the solutions allocate resource for some relays, which are not useful for the system. This is because the subgradient search in the solutions can reduce the duality gap of some bad relays. Through improving the convergence condition of the subgradient search, such types of problems can be avoided. Furthermore, although SC-FDMA is the recommended transmission technique for the LTE uplink, OFDMA is considered for this scheduling scheme. Perhaps, the idea of designing suboptimal algorithm in Chapter 6 can be used if SC-FDMA is taken as the transmission technique. The deployed relays in this problem are of FD functionality, and hence they do not need buffer to store incoming packets. This makes the deployment of relays cheaper and easier.

We have seen in the problem of Chapter 3 that it is difficult for a relay to decide which users to serve under the power constraint even when the assignments of RBs are known. Hence, in Chapter 4, we have proposed a joint sources and relay power allocation scheme for an AF relayed network. We have chosen to use AF relay here because of its simplicity. Existing solutions of this problem ignores the direct link SNR in their problem formulation. Therefore, their schemes are unable to provide correct solution when the direct link SNR of the sources are better than their relayed link ones. Noticing this, we put the direct link SNR back in our formulation, and have solved this using GP-based optimization technique. Having noticed the high computational complexity of the optimal solution, we also have proposed 2 suboptimal solutions of this problem taking greedy and fair nature of the sources into account. Through extensive simulation, we have shown that the proposed solutions correctly allocate power among the sources and relay even when their direct link SNR are better than their relayed link ones. Moreover, the suboptimal solutions achieve very close performance of the optimal one.

We have observed the selfless nature of the users in the solution of the problem in Chapter 4. However, in the real world, the users in the network always want to earn some benefits while sacrificing their resource. To model such behavior of the users, we have proposed a game theoretical solution of this problem in Chapter 5.

Except at the beginning of the game, the nature of the game theoretical solution is distributed. Therefore, it requires relatively less signaling overhead as it is supposed to be. Through extensive simulation, we have shown several properties of this solution. Moreover, we have proved that the proposed solution can achieve comparable performance with the centralized optimal solution under certain convergence condition. The cost functions designed for this solution ignored the log term of the rate functions. The solution would be more accurate if the log term would be incorporated in the cost functions.

In Chapter 6, we have proposed an uplink scheduling and resource allocation scheme for heterogeneous traffic networks. To the best of our knowledge, our proposed scheme is complete for such systems in terms of meeting conflicting requirements of various traffic and taking all standard specific detailed constraints. Besides, the conflicting requirements of different users in heterogeneous traffic networks, the scheme includes a constraint that allows the service provider to keep granular resource utilization level in some prescribed range. The optimal solution of this problem is computationally intensive because of the slow subgradient search. Hence, from the guiding principle of this solution, we have proposed a suboptimal solution with relatively lower complexity. Besides showing the performance improvement achieved by our scheduling scheme comparing with other existing works, we have shown that proposed scheme offers a reasonable tradeoff between ensuring fairness to end users and ensuring effective resource utilization, a consideration of some importance to the network operators. In the simulation, one of the assumptions is that one user holds only one type of traffic. However, our scheduling scheme is general enough to handle the scenario when the users in the network host different types of traffic. Although multi-cell interference is an important issue to consider in the scheduling solution, we have ignored this issue while formulating our scheduling problem.

Although it is apparent that resource allocation in the uplink system can be obtained in a distributed manner, the RBs in the network are global and belonged to the network. Hence, the solutions provided for all problems in this thesis are centralized in nature. Even the game theoretical solution in Chapter 5 is not fully distributed. Main advantage of a centralized solution is that, theoretically, optimal solution can be found. However, centralized solution features high computation and communication overhead, lack of scalability and slow responsiveness. Actual cost of the centralized solution comes from the signaling network. The centralized entity requires a reliable signaling network to gather the information of the entire network. In order to ensure reliable data transmission, the signaling network may need to use protocols with high overhead, such as TCP/IP.

7.2 Future Work

We have listed a set of future works evolved from the outcome of the research work in this thesis.

7.2.1 Self Configurable Uplink Scheduling and Resource Allocation of DF Relayed Systems

One of the assumptions of the problem in Chapter 2 is that the routing information of the entire network is pre-configured. That means, all users know whom to transmit, as well as the relays know which users are connected to them and whom to transmit. Under this setting, it may happen that the users are connected to the bad relays, whereas there are better relays to route. In such circumstances, the performance would be better if the solution schemes would allow the users in selecting helper relays on the fly. Hence, one of the future works is to design a self configurable uplink scheduling scheme of such systems.

7.2.2 Joint Multi-Source and Multi-Relay Power Allocation for AF Relayed Network

In Chapter 3, we have briefly described a joint source and relay power allocation problem on the assumption that one single source is served by one relay. The solution we have given is to maximize the sum throughput of the system. Two more variants of this problem can be

1. Allocate power among the sources and relay in a way that maximizes the minimum source SNR.

2. Allocate power among the sources and relay in a way that minimizes the sum source power.

Furthermore, we can enhance our problem in such a way that instead of single relay, each source can be assisted by multiple relays. Under this network setup, we can formulate the power allocation problem. Similar idea has been discussed in [47], however their problem formulation totally skipped the SNR due to the direct link. Incorporating that factor, we can reformulate the problem and come up with better results.

7.2.3 Game Theoretical Solution of Joint Sources and Relay Power Allocation for AF Relayed Network under Imperfect CSI

One of the assumptions of our work in Chapter 4 is that all entities in the network obtain perfect CSI through the channel estimation and feedback method. However, in practice, this assumption is not realistic and the nodes in the network may not obtain correct CSI. One of the future directions related to this work is to investigate the impact of imperfect CSI on the network due to the channel fluctuations and estimation errors. Bayesian game [138] is very appropriate to capture such condition in the problem.

7.2.4 Uplink Scheduling and Resource Allocation for Multi-Cell Heterogeneous Traffic Networks

One of the drawbacks in the work of Chapter 5 is that the scheduling scheme has not taken interference into account from the neighboring cells. However, this is practical that a cell has many neighboring cells which impose interference on the transmission of the users especially at the edge of that cell. Furthermore, the standard of LTE recommends to have several congested small cells which impose interference on the users of the macro cell. Therefore, one future work in this context is to design an uplink scheduler for heterogeneous traffic networks taking multi-cell interference into account.

Bibliography

[1] G. Intelligence, "Understanding 5G: Perspectives on future technological advancements in mobile," Dec 2014. [Online]. Available: www.gsmaintelligence.com

[2] B. Hanta, "SC-FDMA and LTE Uplink Physical Layer Design," in *Seminar LTE: Der Mobilfunk der Zukunft, University of Erlangen-Nuremberg, LMK*, 2009.

[3] T. S. 3GPP, "Group services and system aspects - policy and charging control architecture," p. 23, Nov 2009.

[4] Ericsson, "Ericsson mobility report: On the pulse of the networked society," June 2014. [Online]. Available: www.ericsson.com

[5] H. Myung, J. Lim, and D. Goodman, "Single carrier FDMA for uplink wireless transmission," *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 30–38, Sept 2006.

[6] 3GPP, "Further advancements for E-UTRA, physical layer aspects," vol. TR 36.813 v0.1.1, p. 23, Nov 2008.

[7] J. Huang, V. G. Subramanian, R. Agrawal, and R. Berry, "Joint scheduling and resource allocation in uplink ofdm systems for broadband wireless access networks," *IEEE J.Sel. A. Commun.*, vol. 27, no. 2, pp. 226–234, 2009.

[8] E. Yaacoub and Z. Dawy, "A game theoretical formulation for proportional fairness in LTE uplink scheduling," in *Proc. IEEE WCNC*, April 2009.

[9] H. Myung, J. Lim, and D. Goodman, "Single carrier FDMA for uplink wireless transmission," *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 30–38, Sept 2006.

[10] H.-L. Chao, C.-K. Chang, and C.-L. Liu, "A novel channel-aware frequency-domain scheduling in LTE uplink," in *Proc. IEEE WCNC*, 2013, pp. 917–922.

[11] X. Wang and S. Konishi, "Optimization formulation of packet scheduling problem in LTE uplink," in *Proc. IEEE VTC-Spring*, 2010, pp. 1–5.

[12] S. N. Marwat, T. Weerawardane, Y. Zaki, C. Goerg, and A. Timm-Giel, "Analysis of radio resource allocation in LTE uplink," *Wirel. Pers.*

Commun., vol. 79, no. 3, pp. 2305–2322, Dec. 2014. [Online]. Available: http://dx.doi.org/10.1007/s11277-014-1986-6

[13] H.-L. Wang, S.-J. Kao, and F.-M. Chang, "An activity selection based single carrier-frequency division multiple access uplink scheduling for LTE networks," in *International Conference on Information Science, Electronics and Electrical Engineering (ISEEE)*, vol. 3, April 2014, pp. 1493–1496.

[14] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath, "Relay architectures for 3GPP LTE-advanced," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, pp. 1–14, 2009.

[15] R. Pabst, B. H. Walke, D. C. Schultz, and et al., "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Magazine*, vol. 42, no. 9, pp. 80–89, September 2004.

[16] L. Le and E. Hossain, "Multihop cellular networks: potential gains, research challenges, and a resource allocation framework," *IEEE Commun. Magazine*, vol. 45, no. 9, pp. 66–73, September 2007.

[17] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, "Resource allocation for OFDMA based relay enhanced cellular networks," in *Proc. IEEE VTC-Spring*, 2007, pp. 3160–3164.

[18] T. Girici, "Joint power, subcarrier and subframe allocation in multihop relay networks," *Int. J. Commun. Syst.*, vol. 22, no. 7, pp. 835–855, Jul. 2009. [Online]. Available: http://dx.doi.org/10.1002/dac.v22:7

[19] T. Wang and L. Vandendorpe, "Sum rate maximized resource allocation in multiple DF relays aided OFDM transmission," *IEEE J. Sel. A. Commun.*, vol. 29, pp. 1559–1571, 2011.

[20] T. Girici, C. Zhu, J. Agre, and A. Ephremides, "Optimal radio resource management in multihop relay networks," in *Proc. WiOpT*, 2008, pp. 443–451.

[21] H. Li, H. Luo, X. Wang, C. Lin, and C. Li, "Fairness-aware resource allocation in OFDMA cooperative relaying network," in *Proc. IEEE ICC*, ser. ICC'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 3750–3754. [Online]. Available: http://dl.acm.org/citation.cfm?id=1817770.1817971

[22] L. Wang, Y. Ji, and F. Liu, "Joint optimization for proportional fairness in OFDMA relay-enhanced cellular networks," in *Proc. IEEE WCNC*, 2010, pp. 1–6.

[23] W.-P. Chang, J.-S. Lin, and K.-T. Feng, "QoS-based resource allocation for relay-enhanced OFDMA networks," in *Proc. IEEE WCNC*, 2011, pp. 321–326.

[24] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, "Fairness-aware radio resource management in downlink OFDMA cellular relay networks," *IEEE Trans. Wirel. Commun.*, vol. 9, no. 5, pp. 1628–1639, 2010.

[25] ——, "Fair resource allocation toward ubiquitous coverage in OFDMA-based cellular relay networks with asymmetric traffic," *IEEE Trans. VEH. TECHNOL.*, vol. 60, no. 5, pp. 2280–2292, 2011.

[26] D. Ng and R. Schober, "Resource allocation and scheduling in multi-cell OFDMA systems with decode-and-forward relaying," *IEEE Trans. Wirel. Commun.*, vol. 10, no. 7, pp. 2246–2258, 2011.

[27] K. Sundaresan and S. Rangarajan, "Efficient algorithms for leveraging spatial reuse in OFDMA relay networks," in *Proc. IEEE INFOCOM*, 2009, pp. 1539–1547.

[28] ——, "Adaptive resource scheduling in wireless OFDMA relay networks," in *Proc. IEEE INFOCOM*, 2012, pp. 1539–1547.

[29] J.-M. Liang, Y.-C. Wang, J.-J. Chen, J.-H. Liu, and Y.-C. Tseng, "Energy-efficient uplink resource allocation for IEEE 802.16j transparent-relay networks," *Comput. Netw.*, vol. 55, no. 16, pp. 3705–3720, Nov. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2011.04.021

[30] D. W. Kifle, O. Bulakci, A. B. Saleh, S. Redana, and F. Granelli, "Joint backhaul co-scheduling and relay cell extension in LTE-ADVANCED networks uplink performance evaluation," in *Proc. European Wireless, 2012*, April 2012, pp. 1–8.

[31] O. Bulakci, A. B. Saleh, S. Redana, B. Raaf, and J. Hamalainen, "Resource sharing in LTE-Advanced relay networks: uplink system performance analysis." *EUR. Trans. Emerg. Telecommun. Technol.*, vol. 24, no. 1, pp. 32–48, 2013.

[32] B.-G. Kim and J.-W. Lee, "Opportunistic resource scheduling for OFDMA networks with network coding at relay stations," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 1, pp. 210–221, 2012.

[33] B. Tang, B. Ye, S. Lu, and S. Guo, "Coding-aware proportional-fair scheduling in OFDMA relay networks," *IEEE Trans Paral. Distr. Sys.*, vol. 24, no. 9, pp. 1727–1740, 2013.

[34] X. S. Xiaoxia Zhang and L.-L. Xie, "Uplink achievable rate and power allocation in cooperative LTE-advanced networks," *IEEE Trans. Veh. Technol.*, pp. 1–12, 2015.

[35] C. Nam, C. Joo, and S. Bahk, "Joint subcarrier assignment and power allocation in full-duplex OFDMA networks," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 6, pp. 3108–3119, June 2015.

[36] B. R. Ho Young Hwang, Hyukjoon Lee and S. Kim, "Joint resource allocation, routing and CAC for uplink OFDMA networks with cooperative relaying," *Wirel. Netw.*, pp. 1–11, Aug 2015.

[37] X. Gong, S. Vorobyov, and C. Tellambura, "Joint bandwidth and power allocation in wireless multi-user decode-and-forward relay networks," in *Proc. IEEE ICASSP*, March 2010, pp. 2498–2501.

[38] L. Han, J. Mu, W. Wang, and B. Zhang, "Optimization of relay placement and power allocation for decode-and-forward cooperative relaying over correlated shadowed fading channels," *EURASIP J. Wirel. Comm. Netw*, vol. 41, 2014.

[39] X. Deng and A. Haimovich, "Power allocation for cooperative relaying in wireless networks," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 994–996, 2005.

[40] T. Quek, M. Win, H. Shin, and M. Chiani, "Optimal power allocation for amplify-and-forward relay networks via conic programming," in *Proc. IEEE ICC*, 2007, pp. 5058–5063.

[41] Y. Zhao, R. Adve, and T. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," *IEEE Trans. Wirel. Commun.*, vol. 6, no. 8, pp. 3114 –3123, august 2007.

[42] W. Su, A. K. Sadek, and K. J. Ray Liu, "Cooperative communication protocols in wireless networks: Performance analysis and optimum power allocation," *Wirel. Pers. Commun.*, vol. 44, no. 2, pp. 181–217, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1007/s11277-007-9359-z

[43] L. Li, Y. Jing, and H. Jafarkhani, "Multisource transmission for wireless relay networks with linear complexity," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2898–2912, 2011.

[44] ——, "Interference cancellation at the relay for multi-user wireless cooperative networks," *IEEE Trans. Wirel. Commun.*, vol. 10, no. 3, pp. 930–939, 2011.

[45] S. Yao and M. Skoglund, "Analog network coding mappings in gaussian multiple-access relay channels," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1973–1983, 2010.

[46] K. Phan, T. Le-Ngoc, S. Vorobyov, and C. Tellambura, "Power allocation in wireless multi-user relay networks," *IEEE Trans. Wirel. Commun.*, vol. 8, no. 5, pp. 2535–2545, 2009.

[47] W. H. Fang, M. J. Deng, and Y. T. Chen, "Joint source and relay power allocation in amplify-and-forward relay networks: a unified geometric programming framework," *IET Commun.*, vol. 5, no. 16, pp. 2301–2309, 2011.

[48] Z. Zhang, H. Liu, and H. Zhang, "Cooperative communications with selfish busy relay's source selection: Whether to cooperate and whom to cooperate with," in *Proc. IEEE WCNC*, April 2014, pp. 2751–2756.

[49] D. Wu, Y. Cai, and M. Guizani, "Auction-based relay power allocation: Pareto optimality, fairness, and convergence," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2249–2259, July 2014.

[50] H. Xiao and S. Ouyang, "Power control game in multisource multirelay cooperative communication systems with a quality-of-service constraint," *IEEE Trans. Intell. Trans. Sys.*, vol. 16, no. 1, pp. 41–50, Feb 2015.

[51] R. Kwan and C. Leung, "A survey of scheduling and interference mitigation in LTE," *JECE*, vol. 2010, pp. 1:1–1:10, Jan 2010. [Online]. Available: http://dx.doi.org/10.1155/2010/273486

[52] F. Capozzi, G. Piro, L. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 2, pp. 678–700, Second 2013.

[53] H. Ramli, R. Basukala, K. Sandrasegaran, and R. Patachaianand, "Performance of well known packet scheduling algorithms in the downlink 3GPP LTE system," in *Proc. IEEE MICC*, 2009, pp. 815–820.

[54] B. Sadiq, R. Madan, and A. Sampath, "Downlink scheduling for multiclass traffic in LTE," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, pp. 14:9–14:9, Mar. 2009. [Online]. Available: http://dx.doi.org/10.1155/2009/510617

[55] G. Monghal, K. Pedersen, I. Kovacs, and P. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution," in *Proc. IEEE VTC-Spring*, 2008, pp. 2532–2536.

[56] S. Saha and R. Quazi, "Priority-coupling-a semi-persistent MAC scheduling scheme for VOIP traffic on 3GPP LTE," in *Proc. IEEE ConTEL*, 2009, pp. 325–329.

[57] M. Lerida, "Adaptive radio resource management for VOIP and data traffic in 3GPP LTE networks," in *KTH Royal Institute of Technology, Stockholm*, 2008.

[58] M. Gidlund and J.-C. Laneri, "Scheduling algorithms for 3GPP long-term evolution systems: From a quality of service perspective," in *Proc. IEEE ISSSTA*, 2008, pp. 114–117.

[59] Y. Zaki, T. Weerawardane, C. Gorg, and A. Timm-Giel, "Multi-QoS-aware fair scheduling for LTE," in *Proc. IEEE VTC-Spring*, 2011, pp. 1–5.

[60] K.-H. Liu, L. Cai, and X. Shen, "Multi-class utility-based scheduling for UWB networks," *IEEE Trans. Veh. Technol.*, 2008.

[61] X. Wang, G. Giannakis, and A. Marques, "A unified approach to QoS-guaranteed scheduling for. channel-adaptive wireless network," *Proc. IEEE*, vol. 95, pp. 2410–2431, 2007.

[62] H. Safa, W. El-Hajj, and K. Tohme, "A QoS-aware uplink scheduling paradigm for LTE networks," in *Proc. IEEE AINA*. Washington, DC, USA: IEEE Computer Society, 2013, pp. 1097–1104. [Online]. Available: http://dx.doi.org/10.1109/AINA.2013.38

[63] S. Kwon and N.-H. Lee, "Uplink QoS scheduling for LTE system," in *Proc. IEEE VTC-Spring*, 2011, pp. 1–5.

[64] H. Ye, G. Lim, L. Cimini, and Z. Tan, "Energy-efficient resource allocation in uplink OFDMA systems under QoS constraints," in *Proc. IEEE MILCOM*, Nov 2013, pp. 424–428.

[65] S. N. K. Marwat, Y. Zaki, C. GÃűrg, T. Weerawardane, and A. Timm-Giel, "Design and performance analysis of bandwidth and QoS aware LTE uplink scheduler in heterogeneous traffic environment." in *Proc. IEEE IWCMC*. IEEE, 2012, pp. 499–504.

[66] S. Cicalo and V. Tralli, "Fair resource allocation with QoS support for the uplink of LTE systems," in *European Conference on Networks and Communications (EuCNC)*, June 2015, pp. 180–184.

[67] R. Musabe and H. Larijani, "Cross-layer scheduling and resource allocation for heterogeneous traffic in 3G LTE," *J. Comp. Netw. Commun.*, pp. 1–13, Aug 2014.

[68] X. T. Samira Niafar and D. H. Tsang, "The optimal user scheduling for LTE-a downlink with heterogeneous traffic types," in *Proc. IEEE QSHINE*, 2014, pp. 1–7.

[69] R. Kwak and J. Cioffi, "Resource-allocation for OFDMA multi-hop relaying downlink systems," in *Proc. IEEE GLOBECOM*, 2007, pp. 3225–3229.

[70] M. Kaneko and P. Popovski, "Adaptive resource allocation in cellular OFDMA system with multiple relay stations," in *Proc. IEEE VTC-Spring*, 2007, pp. 3026–3030.

[71] O. Oyman, “OFDM2A: A centralized resource allocation policy for cellular multi-hop networks,” in *Asilomar Conference on Signals, Systems and Computers, ACSSC*, 2006, pp. 656–660.

[72] C. Muller, A. Klein, F. Wegner, M. Kuipers, and B. Raaf, “Dynamic subcarrier, bit and power allocation in OFDMA-based relay networks,” in *12th International OFDM Workshop*, 2007, pp. 656–660.

[73] L. You, M. Song, and J. Song, “Cross-layer optimization for fairness in OFDMA cellular networks with fixed relays,” in *Proc. IEEE GLOBECOM*, 2008, pp. 1–6.

[74] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, “Resource scheduling for OFDMA/TDD based relay enhanced cellular networks,” in *Proc. IEEE WCNC*, 2007, pp. 1544–1548.

[75] M. Shaqfeh and H. Alnuweiri, “Joint power and resource allocation for block-fading relay-assisted broadcast channels,” *IEEE Trans. Wirel. Commun.*, vol. 10, no. 6, pp. 1904–1913, 2011.

[76] V. Venkatasubramanian and T. Haustein, “A novel scheduling framework for QoS-aware OFDMA resource allocation in a network with small relay cells and macro users.” *EURASIP J. Wirel. Commun. Netw.*, vol. 2012, p. 309, 2012. [Online]. Available: http://dblp.uni-trier.de/db/journals/ejwcn/ejwcn2012.html

[77] K. Sundaresan and S. Rangarajan, “On exploiting diversity and spatial reuse in relay-enabled wireless networks,” in *Proc. ACM MobiHoc*, ser. MobiHoc ’08. New York, NY, USA: ACM, 2008, pp. 13–22. [Online]. Available: http://doi.acm.org/10.1145/1374618.1374622

[78] K.-D. Lee and V. C. Leung, “Fair allocation of subcarrier and power in an OFDMA wireless mesh network,” *IEEE J. Sel. A. Commun.*, vol. 24, no. 11, pp. 2051–2060, Nov. 2006. [Online]. Available: http://dx.doi.org/10.1109/JSAC.2006.881628

[79] L. You, P. Wu, M. Song, J. Song, and Y. Zhang, “Cross-layer optimisation for uplink transmission in OFDMA cellular networks with fixed relays.” *Europ. Trans. Telecommun.*, vol. 22, no. 6, pp. 296–314, 2011.

[80] F. Carmichael, “A guide to game theory,” Dec 2004.

[81] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, “Downlink scheduling and resource allocation for OFDM systems,” *IEEE Trans. Wirel. Commun.*, vol. 8, no. 1, pp. 288 –296, jan. 2009.

[82] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Trans. on Commun.*, vol. 47, no. 6, pp. 884–895, 1999.

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

[84] J. Laneman and G. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. The.*, vol. 49, no. 10, pp. 2415 – 2425, oct. 2003.

[85] T. Wang and G. Giannakis, "Complex field network coding for multiuser cooperative communications," *IEEE J. Sel. A. in Commun.*, vol. 26, no. 3, pp. 561–571, 2008.

[86] G. Sidhu and F. Gao, "Resource allocation for relay aided uplink multiuser OFDMA system," in *Proc. IEEE WCNC*, 2010, pp. 1–5.

[87] Y. Jing and H. Jafarkhani, "Single and multiple relay selection schemes and their achievable diversity orders," *IEEE Trans. Wirel. Commun.*, vol. 8, no. 3, pp. 1414–1423, 2009.

[88] G. Amarasuriya, M. Ardakani, and C. Tellambura, "Adaptive multiple relay selection scheme for cooperative wireless networks," in *Proc. IEEE WCNC*, 2010, pp. 1–6.

[89] H.-S. Chen, W.-H. Fang, and Y.-T. Chen, "Relaying through distributed gabba space-time coded amplify-and-forward cooperative networks with two-stage power allocation," in *Proc. IEEE VTC-Spring*, 2010, pp. 1–5.

[90] M. Ju and I.-M. Kim, "Joint relay selection and opportunistic source selection in bidirectional cooperative diversity networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2885–2897, 2010.

[91] M. Chiang, "Geometric programming for communication systems," *Commun. Inf. The.*, vol. 2, no. 1/2, pp. 1–154, Jul. 2005. [Online]. Available: http://dx.doi.org/10.1516/0100000005

[92] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optimization and Engineering*, vol. 8, no. 1, pp. 67–127, 2007.

[93] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Trans. Netw.*, vol. 13, pp. 636–647, June 2005.

[94] W. Saad, Z. Han, M. Debbah, and A. Hjørungnes, "A distributed coalition formation framework for fair user cooperation in wireless networks," *IEEE Trans. Wirel. Commun.*, vol. 8, no. 9, pp. 4580–4593, Sep. 2009. [Online]. Available: http://dx.doi.org/10.1109/TWC.2009.080522

[95] Z. Zhang, J. Shi, H.-H. Chen, M. Guizani, and P. Qiu, "A cooperation strategy based on nash bargaining solution in cooperative relay networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2570–2577, 2008.

[96] G. Zhang, H. Zhang, L. Zhao, W. Wang, and L. Cong, "Fair resource sharing for cooperative relay networks using nash bargaining solutions," *IEEE Commun. Lett.*, vol. 13, no. 6, pp. 381–383, 2009.

[97] L. Chen, L. Libman, and J. Leneutre, "Conflicts and incentives in wireless cooperative relaying: A distributed market pricing framework," *IEEE Trans. P. Dist. Sys.*, vol. 22, no. 5, pp. 758–772, 2011.

[98] X. Fafoutis and V. Siris, "Performance incentives for cooperation between wireless mesh network operators," in *Proc. IEEE INFOCOM Workshops*, 2010, pp. 1–6.

[99] Y. Shen, G. Feng, B. Yang, and X. Guan, "Fair resource allocation and admission control in wireless multiuser amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1383–1397, 2012.

[100] B. Wang, Z. Han, and K. J. R. Liu, "Distributed relay selection and power control for multiuser cooperative communication networks using stackelberg game," *IEEE Trans. Mob. Comp.*, vol. 8, no. 7, pp. 975–990, 2009.

[101] S. Ren and M. Van Der Schaar, "Pricing and distributed power control in wireless relay networks," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2913–2926, 2011.

[102] Q. Cao, H. V. Zhao, and Y. Jing, "Power allocation and pricing in multiuser relay networks using stackelberg and bargaining games," *CoRR*, vol. abs/1201.3056, 2012.

[103] Q. Cao, Y. Jing, and H. Zhao, "Power allocation in multi-user wireless relay networks through bargaining," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 6, pp. 2870–2882, 2013.

[104] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. A. Commun.*, vol. 13, no. 7, pp. 1341–1347, 1995.

[105] S. Sun and Y. Jing, "Channel training design in amplify-and-forward mimo relay networks," *IEEE Trans. Wirel. Commun.*, vol. 10, no. 10, pp. 3380–3391, October 2011.

[106] ——, "Training and decoding for cooperative network with multiple relays and receive antennas," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1534–1544, June 2012.

[107] F. P. Kelly, A. K. Maulloo, , and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[108] R. Madan, S. Boyd, and S. Lall, "Fast algorithms for resource allocation in wireless cellular networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 973–984, 2010.

[109] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, "Dynamic packet scheduling performance in utra long term evolution downlink," in *Proc. IEEE ISWPC*, 2008, pp. 308–313.

[110] D. Tse and P. Viswanath, "Fundamentals of wireless communication," in *Cambridge University Press; Edition 1*, 2005.

[111] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Downlink scheduling and resource allocation for OFDM systems," *IEEE Trans. Wirel. Commun.*, vol. 8, no. 1, pp. 288–296, 2009.

[112] R. Kwan, C. Leung, and J. Zhang, "Proportional fair multiuser scheduling in LTE," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 461–464, 2009.

[113] F. Calabrese, C. Rosa, K. Pedersen, and P. Mogensen, "Performance of proportional fair frequency and time domain scheduling in LTE uplink," in *European Wireless Conference (EW)*, 2009, pp. 271–275.

[114] N. Chen and S. Jordan, "Throughput in processor-sharing queues," *IEEE Trans. Aut. Contr.*, vol. 52, no. 2, pp. 299–305, 2007.

[115] ——, "Downlink scheduling with probabilistic guarantees on short-term average throughputs," in *Proc. IEEE WCNC*, 2008, pp. 1865–1870.

[116] D. Skoutas and A. Rouskas, "Scheduling with QoS provisioning in mobile broadband wireless systems," in *European Wireless Conference (EW)*, 2010, pp. 422–428.

[117] M. Andrews, "CDMA data QoS scheduling on the forward link with variable channel conditions," *Bell Laboratories*, Apr 2000.

[118] G. Song and G. Li, "Cross-layer optimization for OFDM wireless networks part i: Theoretical framework," *IEEE Trans. Wirel. Commun.*, vol. 4, no. 2, March 2005.

[119] V. J. Venkataramanan and X. Lin, "On wireless scheduling algorithms for minimizing the queue-overflow probability," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 788–801, 2010.

[120] A. L. Stolyar, "Large deviations of queues sharing a randomly time-varying server," *Queueing Syst. The. Appl.*, vol. 59, no. 1, pp. 1–35, May 2008. [Online]. Available: http://dx.doi.org/10.1007/s11134-008-9072-y

[121] B. Sadiq and G. de Veciana, "Large deviation sum-queue optimality of a radial sum-rate monotone opportunistic scheduler," *CoRR*, vol. abs/0906.4597, 2009.

[122] R. Basukala, H. Mohd Ramli, and K. Sandrasegaran, "Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system," in *First Asian Himalayas International Conference on Internet*, 2009, pp. 1–5.

[123] G. Piro, L. Grieco, G. Boggia, R. Fortuna, and P. Camarda, "Two-level downlink scheduling for real-time multimedia services in LTE networks," *IEEE Trans. Multim.*, vol. 13, no. 5, pp. 1052–1065, 2011.

[124] W. K. Lai and C.-L. Tang, "QoS-aware downlink packet scheduling for LTE networks," *Comp. Netw.*, no. 0, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128613000352

[125] M. Iturralde, A. Wei, T. Ali Yahiya, and A. L. Beylot, "Resource allocation for real time services using cooperative game theory and a virtual token mechanism in LTE networks," in *Proc. IEEE CCNC*, 2012, pp. 879–883.

[126] M. Wemersson, S. Wanstedt, and P. Synnergren, "Effects of QoS scheduling strategies on performance of mixed services over LTE," in *Proc. IEEE PIMRC*, 2007, pp. 1–5.

[127] H. Lei, M. Yu, A. Zhao, Y. Chang, and D. Yang, "Adaptive connection admission control algorithm for LTE systems," in *Proc. IEEE VTC-Spring*, 2008, pp. 2336–2340.

[128] I. Wong, O. Oteri, and W. Mccoy, "Optimal resource allocation in uplink SC-FDMA systems," *IEEE Trans. Wirel. Commun.*, vol. 8, no. 5, pp. 2161–2165, 2009.

[129] F. Liu, X. She, L. Chen, and H. Otsuka, "Improved recursive maximum expansion scheduling algorithms for uplink single carrier FDMA system," in *Proc. IEEE VTC-Spring*, 2010, pp. 1–5.

[130] B. Al-Manthari, H. Hassanein, N. A. Ali, and N. Nasser, "Fair class-based downlink scheduling with revenue considerations in next generation broadband wireless access systems," *IEEE Trans. Mob. Comp.*, vol. 8, pp. 721–734, 2009.

[131] D. P. Bertsekas, *Nonlinear Programming.* Belmont, MA: Athena Scientific, 2003.

[132] L. Huawi Technologies Co., "eRAN scheduling feature parameter description," 2013. [Online]. Available: http://www.huawei.com

[133] M. Mojtahed and S. Xirasagar, "Quality of service over LTE networks," 2013. [Online]. Available: www.lsi.com

[134] V. Garg, "Wireless communication & networking," June 2007.

[135] K. YONG-SEOK, "Capacity of VoIP over HSDPA with frame bundling," *IEICE Trans. Commun.*, vol. E89-B, pp. 3450–3453, 2006.

[136] J. Puttonen, T. Henttonen, N. Kolehmainen, K. Aschan, M. Moisio, and P. Kela, "Voice-over-ip performance in utra long term evolution downlink," in *Proc. IEEE VTC-Spring*, 2008, pp. 2502–2506.

[137] J. Blumenstein, J. C. Ikuno, J. Prokopec, and M. Rupp, "Simulating the long term evolution uplink physical layer," in *53rd International Symposium ELMAR*, Zadar, Croatia, September 2011.

[138] D. Fudenberg and J. Tirole, *Game Theory.* Cambridge, MA: The MIT Press, 1991.

Appendix A

Proof of the Properties for I(p)

- *Positivity:* $\mathbf{I}(\mathbf{p}) > 0$. By *Property 5.2*, $\frac{\partial E^*_{s_i}(p_i)}{\partial p_i} < 0$. Moreover, because of practicality, $c_i > 0$, and according to *Lemma 5.1*, $E^*_{s_i}(p_i) \geq 0$. And, following the definition in Equation 5.16, $I_i(p_i) \geq c_i > 0$. Therefore, in this price update process, source $s_i, \forall s_i \in L^1_s$ starts increasing its price from c_i.

- *Scalability:* $\forall \alpha > 1$, $\alpha \mathbf{I}(\mathbf{p}) > \mathbf{I}(\alpha \mathbf{p})$. Subtracting $\mathbf{I}(\alpha \mathbf{p})$ from $\alpha \mathbf{I}(\mathbf{p})$ for source s_i, we have

$$\alpha I_i(p_i) - I_i(\alpha p_i) = (\alpha - 1)c_i + \left[\frac{E^*_{s_i}(p_i)}{\partial E^*_{s_i}(\alpha p_i)/\partial p_i} - \frac{\alpha E^*_{s_i}(p_i)}{\partial E^*_{s_i}(p_i)/\partial p_i}\right]. \tag{A.1}$$

Since $\alpha > 1$, $(\alpha - 1)c_i > 0$. Then, the problem reduces to proving $\frac{E^*_{s_i}(\alpha p_i)}{\partial E^*_{s_i}(\alpha p_i)/\partial p_i} > \frac{\alpha E^*_{s_i}(p_i)}{\partial E^*_{s_i}(p_i)/\partial p_i}$. Now,

$$\frac{E^*_{s_i}(\alpha p_i)}{\partial E^*_{s_i}(\alpha p_i)/\partial p_i} = -2p_i + \frac{2c_i}{\alpha} + \frac{2\sigma\sqrt{B_{s_i}}}{\alpha\sqrt{aE'_{r_i}G_{s_i,r}G_{r,d}}} \left(\alpha p_i - \frac{aG_{s_i,d}}{\sigma^2}\right)^{3/2}, \tag{A.2}$$

$$\frac{\alpha E^*_{s_i}(p_i)}{\partial E^*_{s_i}(p_i)/\partial p_i} = -2\alpha p_i + 2\alpha c_i + \frac{2\alpha\sigma\sqrt{B_{s_i}}}{\sqrt{aE'_{r_i}G_{s_i,r}G_{r,d}}} \left(p_i - \frac{aG_{s_i,d}}{\sigma^2}\right)^{3/2}. \tag{A.3}$$

For the second part of Equation A.1 being > 0, p_i should satisfy $p_i > \frac{\frac{aG_{s_i,d}}{\sigma^2}(\alpha-1/\alpha)}{\alpha-1}$, and $p_i > \frac{aG_{s_i,d}}{\sigma^2}\frac{1/\alpha-\sqrt[3]{\alpha}}{1-\sqrt[3]{\alpha}}$. Or, in other way, $p_i > x\frac{aG_{s_i,d}}{\sigma^2}$, where $x \in \mathbb{R}$. However, x grows slowly with the increasing value of α. Since $c_i > \frac{aG_{s_i,d}}{\sigma^2}$, and $p_i > c_i$ for Case 1, the price update function of source s_i satisfies this property.

- *Monotonicity:* If $\mathbf{p} \geq \mathbf{p}'$, then $\mathbf{I}(\mathbf{p}) \geq \mathbf{I}(\mathbf{p}')$. To satisfy this property, it is sufficient to prove $\frac{\partial \mathbf{I}(\mathbf{p})}{\partial \mathbf{p}} \geq 0$. Hence, for source s_i, we have

$$\frac{\partial I_i(p_i)}{\partial p_i} = 2\left[1 - \frac{3\sigma\sqrt{B_{s_i}}}{2\sqrt{aE'_{r_i}G_{s_i,r}G_{r,d}}}\sqrt{p_i - \frac{aG_{s_i,d}}{\sigma^2}}\right]. \tag{A.4}$$

For $\frac{\partial I_i(p_i)}{\partial p_i}$ being ≥ 0, p_i should be $\leq \frac{aG_{s_i,d}}{\sigma^2} + \frac{4aE'_{r_i}G_{s_i,r}G_{r,d}}{9\sigma^2 B_{s_i}}$. It is apparent that upper bound of p_i for the monotonicity property is close to p_i^{ub} in Case 1 of *Lemma 5.1*.

Appendix B

Proof of the Properties for I(λ)

- *Positivity:* $I(\lambda) \geq 0$. By *Property 5.5*, $\frac{\partial E_{r_i}(\lambda)}{\partial \lambda} < 0, \forall s_i \in L_s^1$. Furthermore, setting the cost c as λ_{lb}, $c > 0$. In Equation 5.20, we observe $E_{r_i}(\lambda) \geq 0, \forall s_i \in L_s^1$. And, according to Equation 5.27, $I(\lambda) \geq c > 0$. Therefore, in this price update process, Network increases price from c.

- *Monotonicity:* if $\lambda \geq \lambda'$, $I(\lambda) \geq I(\lambda')$. To satisfy this property, it is enough to prove $\frac{\partial I(\lambda)}{\partial \lambda} \geq 0$. So,

$$I(\lambda) = c + 2\left[\lambda - \frac{\sum_{s_i \in L_s^1} B_{r_i}}{\sum_{s_i \in L_s^1} \sqrt{\frac{\eta B_{r_i} E^*_{s_i} G_{s_i,r} G_{r,d}}{\sigma^2}}} \lambda^{3/2}\right], \tag{B.1}$$

$$\frac{\partial I(\lambda)}{\partial \lambda} = 2\left[1 - \frac{3\sum_{s_i \in L_s^1} B_{r_i}}{2\sum_{s_i \in L_s^1} \sqrt{\frac{\eta B_{r_i} E^*_{s_i} G_{s_i,r} G_{r,d}}{\sigma^2}}} \lambda^{1/2}\right]. \tag{B.2}$$

For $\frac{\partial I(\lambda)}{\partial \lambda} \geq 0$, it must satisfy that $\lambda \leq \frac{4}{9\sigma^2}\left(\frac{\sum_{s_i \in L_s^1} \sqrt{\eta B_{r_i} E^*_{s_i} G_{s_i,r} G_{r,d}}}{\sum_{s_i \in L_s^1} B_{r_i}}\right)^2$, the right hand side of which is close to λ_{ub} of *Lemma 5.2*.

- *Scalability:* For all $\alpha \geq 1$, $\alpha I(\lambda) \geq I(\alpha\lambda)$. We have

$$\alpha I(\lambda) = \alpha c + 2\alpha\lambda - 2\alpha \frac{\sum_{s_i \in L_s^1} B_{r_i}}{\sum_{s_i \in L_s^1} \sqrt{\frac{\eta B_{r_i} E^*_{s_i} G_{s_i,r} G_{r,d}}{\sigma^2}}} \lambda^{3/2}, \tag{B.3}$$

$$I(\alpha\lambda) = c + 2\alpha\lambda - 2\frac{\sum_{s_i \in L_s^1} B_{r_i}}{\sum_{s_i \in L_s^1} \sqrt{\frac{\eta B_{r_i} E_{s_i}^* G_{s_i,r} G_{r,d}}{\sigma^2}}}\alpha\lambda^{3/2}. \tag{B.4}$$

Since $\alpha > 1$, and c is positive, $(\alpha - 1)c$ is always > 0. Furthermore, for the second part of $\alpha I(\lambda) - I(\alpha\lambda)$ being positive, it must satisfy that $\alpha^{3/2} \geq \alpha$, which is true for all $\alpha \geq 1$

Appendix C

Additional Research Work

Some research works that are not directly related to this dissertation but have been published during my time as a PhD student at UBC are as follows.

- Li Zhou, Chunsheng Zhu, Rukhsana Ruby, Xiaofei Wang, Xioating Ji and Jibo Wei, "QoS-Aware Energy-Efficient Resource Allocation in HetNets," *accepted for International Journal of Communication Systems*, 2014, p. 1–19.
- Xuan Dong, Shaohe Lv, Rukhsana Ruby, Yong Lu, Xiaodong Wang, XingMing Zhou and Victor C.M. Leung, "Virtual Frame Aggregation: Cluster Contention Based Channel Access in Multicarrier Wireless Networks," *IEEE ICC*, 2015, p. 1–7.
- Li Zhou, Rukhsana Ruby, Haitao Zhao, Xiaoting Ji, Jibo Wei and Victor C.M. Leung, "A Graph-based Resource Allocation Scheme with Interference Coordination in Small Cell Networks," *IEEE GLOBECOM Workshops*, 2014, p. 1–6.
- Rukhsana Ruby, Saad Mahboob and David G. Michelson, "Optimal Configuration of Distributed MIMO Antennas in Underground Tunnels," *IEEE APS/URSI*, 2014: 65–66.
- Saad Mahboob, Rukhsana Ruby, and Victor C.M. Leung, "Transmit Antenna Selection in a Very-Large MIMO System using Convex Optimization," *IEEE BWCCA*, 2012: 228–233.

- Rukhsana Ruby, Victor C.M. Leung and John Sydor, "Optimal Transmission Strategy for a Secondary User in IEEE 802.11 Based Networks," *IEEE PIMRC*, 2012: 1243–1248.

- Rukhsana Ruby, Salim Hanna, John Sydor and Victor C.M. Leung, "Interference Sensing using CORAL Cognitive Radio Platforms," *Chinacom*, 2011: 949–954.

- Rukhsana Ruby, and Victor C.M. Leung, "Performance Analysis of a Heterogeneous Traffic Scheduler using Large Deviation Principle," *to be submitted to Computer Communications*, 2015.