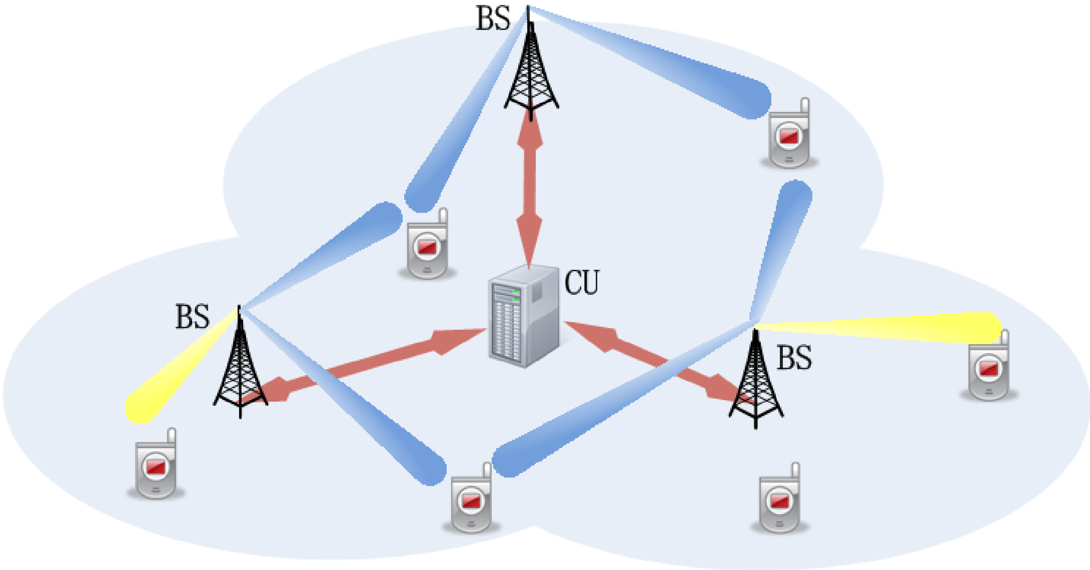
CHALMERS



Utility-Based Scheduling for Multi-Service in Coordinated Multi-Point Networks

Master of Science Thesis

BINRU HUANG

Department of Signals and Systems
Communication Systems Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2012

Report No. EX006/2012

MASTER OF SCIENCE THESIS 2012

Utility-Based Scheduling for Multi-Service in Coordinated Multi-Point Networks

Binru Huang

Supervisor: Jingya Li
Examiner: Tommy Svensson

Department of Signals and Systems
Communication Systems Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2012

Utility-Based Scheduling for Multi-Service in Coordinated Multi-Point Networks

Master of Science Thesis 2012

Department of Signals and Systems
Communication Systems Group
Chalmers University of Technology
SE-41296 Gothenburg
Sweden

Tel. +46-(0)31 772 1000

Department of Signals and Systems
Gothenburg, Sweden 2012

Abstract

Coordinated Multi-point (CoMP) joint transmission has been identified as one of the key techniques to mitigate inter-cell interference (ICI), improve the spectral efficiency and create a more uniform service experience in cellular systems. In CoMP systems, base stations cooperate in jointly transmitting signals to multiple users in the downlink, or receiving signals from multiple users in the uplink, respectively. In this thesis, the issues about user scheduling in downlink packet transmission with multiple traffic models for CoMP networks is addressed. Two different traffic patterns, e.g., voice over IP (VoIP) and best-effort (BE) are considered and modeled. Utility-based coordinated scheduling algorithms are proposed to achieve the maximum sum utility in the mixed-traffic CoMP network. We first consider a CoMP system with flat fading channel and a single cluster, which consists of three neighboring base station sectors (BSSs). To evaluate the performance of the proposed algorithm, the traditional scheduling schemes without CoMP transmission and diverse QoS provisioning are also investigated. By comparison, the simulation results show that the proposed algorithm increases the cell-edge BE efficiency by greater than 45%, while better satisfying the QoS requirements for VoIP with greater than 98% decrease in call outage.

Based on the formentioned utility-based joint scheduling algorithm, we extended our work to multi-cell multi-service orthogonal frequency division multiple access (OFDMA) scenarios. In order to solve the scheduling problem in the multi-cell multi-service OFDMA system with CoMP joint transmission, the ant colony optimization (ACO), a typical algorithm of metaheuristic methods, is introduced in this thesis. A two-step utility-based joint scheduling algorithm is proposed: firstly on each subchannel users are jointly selected based on the ACO scheme, and then all the subchannels are allocated for each group of coordinated BSSs by using a greedy algorithm. We compare the proposed algorithm with the optimal algorithm and the greedy user selection (GUS) based scheme. The simulation results illustrate that the performance of our proposed algorithm well approaches that of the optimal algorithm. It is also shown that our algorithm ensures that more than 95% of VoIP users are satisfied with packet drop ration less than 2%, compared to 78% by the GUS based algorithm. Meanwhile, 95% of BE users are satisfied with average cell-edge data rate greater than 200kbps by using either our proposed algorithm or the GUS based algorithm.

Keywords: coordinated multi-point (CoMP), OFDMA, multiple services, quality of service (QoS), scheduling and resource allocation.

Acknowledgements

First of all, I wish to express my deepest gratitude towards those people who made this master thesis possible, including Professor Tommy Svensson, my supervisor Jingya Li, and the PhD student Yutao Sui. A special thank to my manager Stefan Landelius at Ericsson AB for offering me a lot of help sponsoring my trip to the *IEEE Vehicular Technology Conference (VTC) 2012 Spring* in Yokohama, Japan. I'm very grateful to have Tommy and Jingya as my supervisors. Tommy has offered me a great help and support all the way during this thesis work. I thank you so much for encouraging and sponsoring me to attend and present our papers at the *IEEE Wireless Personal Multimedia Communication (WPMC) Conference 2011* in Brest, France, as well as at the *VTC 2012 Spring Conference*. It is a precious and memorable experience for me. Jingya's constructive and brilliant comments on my thesis have been well appreciated. Your knowledge, guidance and encouragement have made a huge difference. I thank you for being kind enough to answer my endless questions and sharing your intriguing perspective on the thesis. I also thank Yutao for his help and support during the summer internship at Chalmers, which well prepared me for this master thesis.

Once again, I would like to thank Tommy and Jingya for offering me such an interesting thesis topics, for guiding and inspiring me all the way through this thesis work, as well as for encouraging and supporting me all the way of my job searching. My sincere gratitude goes to Tommy for your helpful and supporting references, which made it possible for me to find such a good job at Ericsson.

For helping me with administrative tasks related to this thesis, I send my gratitude to Madeleine Persson, Agneta Kinnander, and Lars Börjesson. I thank Madeleine for offering a lot of helps in finding me a nice room for my summer internship and master thesis, as well as for your helps with the application for my visa extension. I would like to thank Lars for your supports in supplying computers for this thesis. I also thank Agneta Kinnander, our department secretary, for your kind help and support during my trip to Brest.

Besides, I am grateful to all the teaching staff and PhD students of the communication system group at Chalmers. Everyone in the group made me feel welcome in the group. I had a great time in our department social day.

My sincerest gratitude goes to my family. I want to thank my family, for their love and support all these years. Finally, I would like to thank my boyfriend Songhe. The joy and constructive suggestions you give me has been invaluable in the writings of this thesis.

Finally, I would like to thank all the people I mentioned here. Through this master thesis I earned a great of joy, achievements, confidences and growth, and of most importance, friendship, all of which are invaluable to me. Thank all of you.

Binru Huang, Gothenburg 2/5/2012

Acronyms

3GPP	3rd Generation Parnership Project
3G	3rd Generation Wireless
ACO	Ant Colony Optimization
AMC	Adaptive Modulation and Coding
AP	Access Point
AWGN	Additive White Gaussian Noise
BE	Best Effort
BER	Bit Error Rate
BLER	Block Error Rate
BMB	Buffer Management Block
BS	Base Station
BSC	Base Station Coordination
BSS	Base Station Sector
CBS	Coordinated Base-station-sector Set
CCU	Cell-Center User
CDMA	Code Division Multiple Access
CEU	Cell-Edge User
CoMP	Coordinated Multi-Point
CFR	Cooperative Frequency Reuse
CSI	Channel State Information
CS/CB	Coordinated Scheduling/Beaforming
CQI	Channel Quality Information
CU	Central Unit

DCS	Dynamic Cell Selection
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
GSM	Global System for Mobile communication
GUS	Greedy User Selection
HOL	Head Of Line
ICI	Inter-Cell Interference
ISI	Inter-Symbol Interference
JFI	Jain's Fairness Index
JP	Joint Processing
LTE	Long Term Evolution
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MS	Mobile Station
MUD	Multi-User Diversity
M-LWDF	Modified-Largest-Weighted-Delay-First
OFDM	Orthogonal Frequency Division Multiple
OFDMA	Orthogonal Frequency Division Multiple Access
PER	Packet Error Rate
PC	Power Control
PC	Packet Classifier
PLR	Packet Loss Rate
PHY	Physical
PS	Packet Scheduler
RB	Resource Block

RT	Real-Time
SINR	Signal to Interference and Noise Ratio
SISO	Single-Input Single-Output
SNR	Signal to Noise Ratio
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UB	Utility-Based
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
VoIP	Voice over Internet Protocol
WCDMA	Wideband Code Division Multiple Access

Contents

PART I

Introduction

1.1 Introduction and Motivation

Future wireless networks will provide access to multiple applications with different requirements of the quality of service (QoS). For example, video telephony, file transfer and web-surfing require high-speed data rate, but voice over IP (VoIP) demands relatively low-rate applications, though critical requirements on latency. With different demands of diverse wireless services, orthogonal frequency division multiple access (OFDMA) has been considered as a leading technique for future wireless networks [1]. Additionally, universal reuse of spectrum is envisioned to improve spectrum efficiency. However, universal frequency reuse meanwhile brings inter-cell interference (ICI), which could significantly degrade the system performance and impair the QoS. Especially at the cell-edge area, the degradation of wireless services due to high levels of ICI could cause significant impairment to the fairness of wireless network systems.

Coordinated Multi-Point (CoMP) transmission, also known as network coordination, is proposed as a promising technique to satisfy the system requirements regarding coverage, cell-edge user throughput and system efficiency [2]. In CoMP systems, a group of base stations (BSs) are connected via a high-speed backbone, and transmit signals coordinately to users so as to mitigate ICI, and hence improve spectral efficiency. In the downlink of CoMP systems, the ICI is significantly mitigated by applying the signals transmitted from other cells to assist the transmission instead of acting as interference by using CoMP joint transmission. Current studies in the literature of CoMP schemes mainly focus on increasing cell-edge throughput and fairness assuming full buffers for all the users [3]-[6]. However, other critical system metrics than spectral efficiency and fairness are not treated like delays, service outage, etc.

On the other hand, to meet diverse QoS requirements imposed by various services, research on mixed-traffic scenarios is receiving more and more attention due to its significance in practical deployment of the current wireless networks [7]-[10]. Priority-based and utility-based scheduling algorithms based on channel-aware and queue-aware systems, with respect to diverse metrics, e.g., system throughput, latency, service outage, are proposed for dynamic systems. These algorithms are shown to better satisfy the strict delay requirements of real-time (RT) traffic while balancing the fairness and spectral efficiency. However, the main limitation is that all these scheduling methods are designed mainly for single-cell mixed-traffic scenarios, without considering joint transmission.

In this thesis, the problem of CoMP joint transmission in mixed-traffic networks is addressed. Two different traffic patterns are taken into account, i.e., best effort (BE) and VoIP traffic. We model two different utility functions presenting the satisfaction level of users based on two different traffic patterns, respectively. A utility-based joint packet scheduling and power control algorithm is proposed by maximizing the aggregate utility to achieve the users' satisfaction in CoMP networks. Binary power control (i.e., the base station either transmits with full power or does not transmit in each time slot) is also exploited to achieve higher system performance. Proportional-fair joint scheduling and two other schemes without joint transmission are also evaluated and compared with our proposed algorithms in the downlink of CoMP networks. To simplify the scenario, we first assume in our simulations a CoMP system that 1) a flat fading channel, and 2) a single cluster scenario, where a cluster is defined as a group of three neighboring base station sectors (BSSs).

Based on the proposed utility-based joint scheduling algorithm, we extended our work to multi-cluster multi-service OFDMA scenarios. However, the proposed utility-based joint scheduling algorithm, based on exhaustive search for the optimal solution, is shown to be unpractical in addressing the joint scheduling problem in the multi-cluster multi-service OFDMA system. The computational complexity of the algorithm, especially when the number of users is large, could be prohibitively high for practical use. In [5], [11] and [12], the combinatorial optimization problem of joint scheduling in coordinated OFDMA networks with full buffer traffic was addressed. In [5] a convex optimization based coordinated subchannel assignment scheme was proposed with the objective of maximizing total cell-edge utility. In [11] a greedy user selection (GUS) based algorithm is proposed, reducing the computational complexity of centralized multi-cell scheduling. In [12] a graph theory based two-phase scheduling scheme was proposed to handle the joint scheduling problem. Besides the traditional algorithms mentioned above, ant colony optimization (ACO) [13] has recently attracted growing attention in the studies of complex combinatorial optimization problems. ACO was proposed in [14] to search for sub-optimal solutions for NP-hard problems, and then has become widely used for channel allocation in wireless networks [15]-[17]. It is illustrated in [14]-[17] that ACO can achieve better performance than other traditional sub-optimal algorithms.

Notice that, [11]-[12] assume only a single service for all the users in the CoMP networks, and in [15]-[17] CoMP joint transmission is not considered. Hence, address the joint scheduling problem in multi-cluster multi-service OFDMA systems with CoMP joint transmission, we employ the ACO for utility-based joint scheduling, and propose a two-step joint scheduling algorithm. Based on ACO, a joint user selection scheme is designed for each subchannel as a first step. Then subchannel assignment is implemented by using a greedy algorithm.

A major contribution of this thesis work is that we propose a utility-based joint scheduling to address the CoMP joint transmission problem in CoMP networks with multiple services. Besides BE users with full buffer, the delay-sensitive VoIP traffic is also considered and modeled. We investigate the performance on the downlink of flat fading channel in single-cluster CoMP system, as well as extend our research to a multi-cluster OFDMA system, and proposed a sub-optimal algorithm by combining ACO and utility-based scheduling.

The remainder of this chapter is organized as follows. In Section 1.2, the background and the state of the art of CoMP system is described. From an economic network point of view, current techniques of multi-user downlink data scheduling and resource allocation, with QoS provisioning over shared fading channels, are also presented in this section. The CoMP system models used in our thesis are described in Section 1.3. Then the traffic models, as well as the queue structure for packet scheduling are presented, respectively. In the end of this section, the ACO system is introduced. The contributions of this thesis are briefly described in Section 1.4, and the limitations of the work in this thesis, as well as the prospect for future work will be discussed in 1.5.

1.2 Background and Related Work

One key characteristic of wireless networks is its time-varying nature. There are several reasons for these variations. Firstly, frequency-selective fading will lead to rapid and random fluctuations in the channel attenuation. The average received signal strength is also affected significantly by shadow fading and distance-dependent path loss. Moreover, the rapid change of ICI will also impact the interference level at the receiver side [18]. Hence, in order to improve system performance and resource utilization in the wireless networks, variations of the quality of the wireless channel must be taken account into and preferably exploited. Hence, the major question of importance is how these variations can be exploited by a scheduling algorithm in wireless networks.

In this section, we will first review some of the basic information theories, followed by the current well-known techniques in wireless networks, including link adaption and channel-dependent scheduling. A brief review of OFDMA networks and resource allocation for OFDMA system will also be addressed, as well as multimedia services, and QoS provisioning utilities from a view of network economics. Finally, the concept of CoMP transmission will be described.

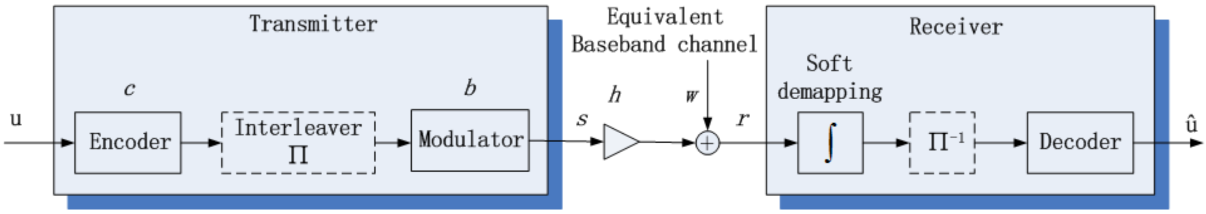


Figure 1.1: Discrete-time system diagram of MCS.

1.2.1 Throughput and Link Adaptation

Link adaptation deals with how to set the transmission parameters of a radio link to handle variations of the radio-link quality, generally including rate and power control. In practice, radio-link rate is controlled by adjusting the modulation scheme and/or the channel coding rate based on the received signal-to-noise-ratio (SNR) to approximate to the channel capacity. For this reason, radio-link rate control is sometimes also referred to as adaptive modulation and coding (AMC). Power control, on the other hand, adjusts the transmit power to compensate for the variations of the radio-link quality to maintain a constant received SNR. It has been shown that rate control is more efficient than power control. However, in practice, rate control is often utilized combined with power control [18].

Throughput and Rate Adaptation

Consider a wireless point-to-point link with one transmitter and one receiver communicating through an additive white Gaussian noise (AWGN) channel, which can be simplified and modeled as in Fig.1.1. In the system, data from upper layer, i.e., medium access control (MAC) layer, is encapsulated into packets of bits, which is denoted by the vector $\mathbf{u}$, and sent to the physical (PHY) layer of the transmitter. The transmit packet bits are then encoded and interleaved. An interleaver Π is not always essential, but usually exists to protect the transmission against burst errors. After encoding and interleaving, the bits are then modulated into symbol stream s before being transmitted over the discrete-time equivalent baseband channel with channel gain h, and AWGN w with variance of σ^2 and zero-mean. The channel output is given as $r = h \cdot s + w$. At the receiver side, the opposite operations are performed and estimated vector $\widehat{\mathbf{u}}$ is acquired in the end. Assume the signal-to-noise ratio (SNR) at the receiver side is γ over the given frequency band with bandwidth B, the Shannon's channel capacity of the equivalent baseband discrete-time channel in Fig.1.1 is defined by [18]

$$C(\gamma) = B \log_2(1 + \gamma). \tag{1.1}$$

However, the Shannon's capacity formula gives the upper bound of the data rate based on ideal channel coding and data modulation schemes. In practice, throughput is defined as the amount of error-free data that is delivered to the upper layers of the communication system after demodulation and decoding at the receiver side. For packet-based transmission, the error probability of received packets are named packet error rate (PER). Generally, PER depends on the selected modulation and coding scheme (MCS) and the instantaneous received SNR. Assume that MCS m with $b^{(m)}$ representing bits carried by per symbol, and $c^{(m)}$ denoting the coding rate. The effective data rate is then given as $R^{(m)} = b^{(m)}c^{(m)}$. With an instantaneous received SNR γ, packets are received with the PER of $\mathrm{PER}^{(m)}(\gamma)$ for MCS m, and thus the expectation of the throughput is expressed by

$$\bar{\eta}^{(m)}(\gamma) = \mathbf{E_t}\{\eta^{(m)}(\gamma)\} = R^{(m)}\left(1 - \mathrm{PER}^{(m)}(\gamma)\right). \tag{1.2}$$

In practice, the PER can be approximated by a step function, i.e.,

$$\mathrm{PER}^{(m)} \simeq \begin{cases} 1, & \text{for } \gamma \leq \gamma^{(m)}; \\ 0, & \text{for } \gamma \geq \gamma^{(m)}. \end{cases} \tag{1.3}$$

where $\gamma^{(m)}$ is a constant for MCS m usually depending on the prescribed requirements of minimum PER.

In order to maximize the system throughput, rate adaption referred to the capability of the system to adapt parameters such as modulation and coding rate to the quality of the wireless link, has been proved to be effective in exploiting the variations of wireless channel states [19]-[21]. By rate adaptation, when channel is poor, the most reliable MCS will be chosen such that the target PER is satisfied; otherwise, when channel is good, the MCS offering the maximum effective transmission rate is used. Especially, in case of packet-based traffic with diverse data rate requirements, there is no need to provide a constant data rate over a radio link. Therefore, by means of adaptive MCS, the system can adjust the data rate to compensate for the varying channel states, and hence in turn achieve an improvement in system throughput and resource utilization.

Power Allocation

Besides MCS adaptation, dynamic transmit-power allocation also receives a lot of concerns in the studies of link adaption. Historically, dynamic transmit-power control has been used in wideband code division multiple access (WCDMA) and cdma2000 to compensate for variations in the instantaneous channel conditions [18]. As the name suggests, by power adaptation the transmit power is allocated depending on the instantaneous channel states, in order to maintain a fixed error probability or, equivalently, a constant received SNR. Therefore, the power allocation strategy compensates SNR by inverting the channel fading [18].

However, the channel inversion scheme mentioned above still achieves low data in poor channel. Moreover, in case of packet-based transmission, there is no need to provide constant data rate. Instead, to achieve the channel capacity, water-filling policy is proved to be the optimal power allocation strategy [22]. The principle of water-filling is to take advantage of good channel conditions, namely when the quality or SNR of a radio link is high, more power will be allocated to the link so as to achieve a higher data rate. When the channel quality is poor, less power is assigned to the link, and if the instantaneous SNR falls below a specific cutoff SNR, the channel is not used [18].

Link adaptation with adaptive MCS and power allocation has been widely used in packet-based transmission for compensation of variable channel states [19]-[21], [22]. However, when channel experiences a continuously deep fading, link adaptation is not able to provide any enhancement in spectral efficiency for a single point-to-point link. Therefore, to enhance the system throughput, the channel diversity among multiple links should also be taken into account.

1.2.2 Resource Allocation in Multi-User SISO System

As we discuss in Subsection 1.2.1, for a point-to-point link simplified as in Fig.1.1, link adaption can exploit the variations of the channel quality, and hence improve the throughput. However, if the channel experiences deep fading, the throughput is still low. Though link adaptation is still an important technique in resource allocation designing on PHY or MAC layers, the principles of resource allocation designs for multi-user downlink have been recently shifted from the perspective of conventional point-to-point link to a multi-user network view. In a multi-user wireless network, different users experience independent channel conditions, resulting in that some users experience better channel conditions than others at a time instant. This phenomenon is known as multi-user diversity (MUD). Therefore, taking advantage of MUD users with better channel conditions can better exploit the channel quality and yield a better performance. Clearly, if CSI is known to the scheduler or the resource allocation unit, it is possible to exploit MUD.

Channel-Dependent MUD Scheduling

Consider the downlink of a centralized multi-user network with K single antenna mobile users, and a single antenna access point (AP). See Fig.1.2 for an example of a multi-user SISO network. A perfect instantaneous feedback of CSI is assumed for AP to implement MUD scheduling. We also Assume a block fading channel such that a channel realization $\mathbf{h} = \{h_1, h_2, \cdots, h_K\}$ is constant in a duration T_s called transmission time interval (TTI), but changes independently every T_s seconds. The power allocation vector is denoted as $\mathbf{p}(\mathbf{h}) = \{p_1, p_2, \cdots, p_K\}$ according to the CSI. In order to maximize system throughput subject to the maximum total power constraint, we have the optimal solution as

$$\begin{aligned} k^* &= \arg\max_{\mathbf{h}} R_k \\ &= \arg\max_{\mathbf{h}} |h_k|^2, \end{aligned} \tag{1.4}$$

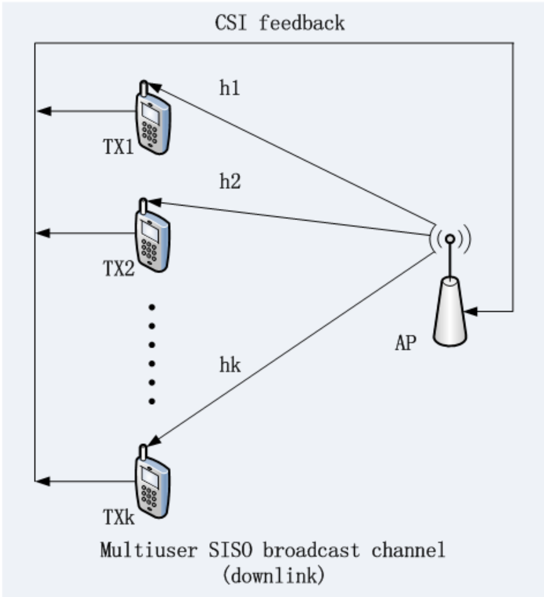


Figure 1.2: Downlink of a multi-user SISO network.

where R_k is the throughput of user k expressed by

$$R_k = \log_2 \left(1 + \frac{p_k |h_k|^2}{\sum_{j \neq k} p_j |h_j|^2 + \sigma^2} \right), \tag{1.5}$$

according to Shannon's channel capacity. σ^2 is the variation of the zero mean Gaussian noise at the receiver side, $\sum_{j \neq k} p_j |h_j|^2$ is the interference caused by other radio links in the network, and the expression

$$\gamma = \frac{p_k |h_k|^2}{\sum_{j \neq k} p_j |h_j|^2 + \sigma^2} \tag{1.6}$$

is referred to as signal-to-interference-and-noise ratio (SINR). Hence, the MUD scheduling problem becomes a problem of allocating all power to the user k^* with the best channel condition. This scheduling algorithm is also referred to as max rate scheduling.

Multi-User Fairness and Proportional Fair

However, for future wireless networks, in addition to throughput, user fairness is also another important index to evaluate a scheduling algorithm. In wire-line networks, fair typically means all the users equally share the resources, i.e., all the users receive the same throughput. However, in the context of wireless networks, different users experience different and independent channel conditions, resulting in vast fluctuations in the average channel qualities for different users. If we still consider the fairness in the perspective of throughput, then max rate scheduling is not a fair scheduling. Take the network in Fig.1.2 for an example. With max rate scheduling, a user experiencing a long-term deep channel attenuation in the network, can suffer from poor service, namely, the user could hardly get scheduled due to its poor channel quality. An alternative to the max rate scheduling algorithm is so-called round robin scheduling, with which users are scheduled orderly. However, this strategy doesn't utilize MUD, so it is neither fair in the sense of providing the same service quality to all the users, and will lead to lower overall system throughput [18].

Therefore, a practical scheduler should satisfy some degree of fairness between users, and also be able to take advantage of fast variations in channel states as much as possible. One example of such a scheduler is the proportional fair scheduler [23]. In this algorithm, the shared resource is assigned to the user k^* according to

$$k^* = \arg\max_k \frac{R_k(\mathbf{h})}{\bar{R}_k}. \tag{1.7}$$

Hence, if a user experience poor channel conditions and is not scheduled for a long time, the user get its priority increased by the low denominator in (1.7), and would more probably get scheduled. From the view of a long term performance, proportional-fair scheduler is a fair scheduler in terms of average throughput. To measure fairness, Jain's fairness index (JFI) is often used [24]. If we define the fairness in the view of fair throughput, then the corresponding throughput-fairness index JFI is given as

$$J = \frac{\left(\sum_{k=1}^{K} \bar{R}_k\right)^2}{K \cdot \sum_{k=1}^{K} \bar{R}_k{}^2}. \tag{1.8}$$

Notice that JFI always lies in $(0, 1]$. A JFI of 1 indicates a throughput-fair scheduling algorithm; on the other hand, a JFI of 0 indicates a throughput-unfair algorithm.

1.2.3 Orthogonal Frequency Division Multiplexing

It is also known that frequency selectivity is another important characteristic of wireless channels, i.e., the channel quality varies in the frequency domain. Therefore, in order to exploit the MUD in the frequency domain, OFDM technique is also addressed in this thesis. Currently, OFDMA is being considered as a promising multiple access method for next generation wireless networks. OFDMA has been adopted as the downlink access technology of 3rd generation parternership project (3GPP) long term evolution (LTE) and LTE-Advanced standards [1], [25].

Basic Principles of OFDM

As we know, in broadband wireless networks, frequency selectivity is a major cause of inter-symbol interferences (ISI) in radio transmission resulting in severe signal distortion.

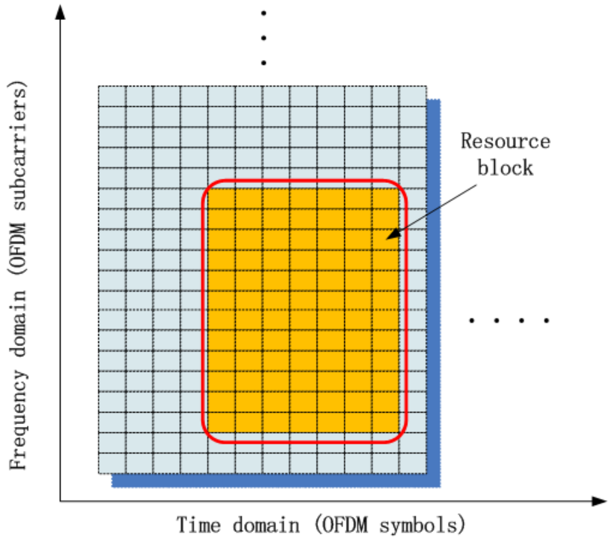


Figure 1.3: OFDM time-frequency grid.

With OFDM, the broadband channel is divided into multiple parallel and relatively flat narrowband subcarriers, and thus OFDM transmission can provide immunity to ISI in transmission.

In practice, a fixed number of OFDM symbols are transmitted in a time slot, and during each time slot, a fixed number of subcarriers are grouped to form a resource block (RB). See Fig.1.3 for the illustration of the time-frequency grid of an OFDM system, as well as OFDM RBs. In general, an RB is the minimum scheduling unit for OFDM resource allocation. Specially, in case of OFDM transmission over highly frequency-selective broadband channel, since multiple users experience diverse and independent channel states on different subchannels, with CSIs OFDM transmission can exploit frequency diversity and MUD gains and achieve a significant enhancement in spectral efficiency.

Resource Allocation for OFDMA

Resource allocation schemes for OFDMA systems have been studied by the authors in [26]-[28]. Proposed approaches mainly focus on the physical layer transmission optimization for OFDMA systems, and can be categorized into two different schemes: (i) minimizing the total transmit power subject to QoS requirements for each user such as data rate, bit error rates, delays, etc. [27]; and (ii) maximizing the data rates subject to various QoS and/or resource constraints, e.g., maximum total transmission power [26], [28]. In [26], the authors present two centralized heuristic algorithms. The first algorithm considers each subcarrier independently and assigns users to each subcarrier with adaptive modulation scheme separately. The second exploits water-filling across all the subcarriers. Both of the two algorithms are designed to maximize data rate subject to the constrains of maximum BER and total transmit power. In [27], a subcarrier allocation algorithm, aimed at minimizing the total transmit power is developed by allowing each user to specify its individual requirements of both bit rate and BER. In [28], a nonconvex dual problem is studied, namely the multi-user spectrum optimization subject to the total transmit power constraint. A low-complexity method to approximate the evaluation of the Lagrangian dual objective is then proposed using subgradient and ellipsoid algorithms.

1.2.4 Multi-Service and QoS Requirements

Future wireless networks are expected to support a wide variety of services and applications such as e-mail, transfer of data files, voice over IP and video conferencing. Therefore, in addition to spectral efficiency, other metrics like e.g., minimum rates for file transfers and delay constraints for VoIP and video streaming also need considerations and should be taken into account in scheduling design for the future networks. Unlike wire-line networks, wireless resources are quite limited and valuable; moreover, with the nature of time-varying of wireless networks, how to design a scheduling algorithm that could achieve high throughput while satisfying diverse QoS requirements has become a major question of importance.

In general, utility is often used to quantify the level of satisfaction and to address decision of resource allocation in networking [29]. For instance, utility function like logarithm function has been widely studied and exploited to address the issues on efficient and fair resource allocation. Similarly, utility theory can be used in communication networks to evaluate the degree to which a network satisfies service requirements of users' applications, rather than in terms of system-centric quantities like throughput, outage probability, packet drop rate, latency, etc.. The basic idea of utility-based scheduling is to allocate the radio resources such as time, frequency band, and power, or the performance criteria like data rate or delay, etc. into the corresponding utility and optimize the utility-based system. Hence, the utility of an application usually depends on the application performance, which in turn depends on the QoS requirements of the system [30], [31].

Traffic Types and QoS requirements

Usually, applications used in the internet are often divided into the following categories based on their different QoS requirements for data rates and latency [7]:

- **Best effort (BE)** requires high data rate, but impose no bound on delay, e.g., web browsing and email.
- **Real-time (RT)** has very strict demands for latency, but requires relatively lower date rate such as VoIP service, and video conferencing.

Therefore, besides throughput delay is also an important factor that has an impact on the system performance.

Recently, mixed-traffic scheduling has received more and more concerns, and a lot of QoS-based opportunistic scheduling algorithms are proposed to address the issues in the trade-off between spectral efficiency and delay requirements in mixed-traffic networks [7]-[10]. One approach is to assign RT services with the strict highest transmit priority among these three classes of traffic, while BE services are assigned the lowest priority, at the price of decrease in the system spectral efficiency [8]. By contrast, in [9] a higher utility is assigned to RT users, compared to BE users, dynamically based on the number of time slots the RT users has waited in the system queue, so as to maintain relatively high throughput while satisfying the delay requirements of RT users. In [7], a unified approach for QoS-guaranteed scheduling is designed to address resource allocation for mixed-traffic networks based on diverse utilities associated with various QoS requirement for different traffic users. In [10], the authors propose utility-based scheduling algorithms that are aware of both channel and queue state information to achieve the maximum aggregate utility in the network, which significantly improve the performance in balancing multi-user diversity and delay.

Transmit Delay and Delay Metrics

In packet-based communications, end-to-end delay is the total time a packet takes from its source until it reaches its destination, where this total delay is the summation of a variety of delay sources in the system [34]. A widely known source of delay is the queuing delay, i.e., the time a packet waits in the system queue for transmission until it arrives at the head of the queue. Another source of delay is the access delay, accounting for the time a user waits until it gets access to the channel. These two types of delay compose the transmit delay, which is defined in this thesis as *transmit delay* = *queuing delay* + *access delay*, characterizing the time interval a packet waits at the transmitter side from the instant it arrives at the tail of the system queue until it is finally transmitted to the channel. Sometimes, the aggregate of queuing delay and the access delay is also generally called as the queuing delay without distinguishing between the two delays. Other delay sources like propagation delay or packet processing delay at the both transmitter and receiver sides, as well as the transmit delay, compose the end-to-end delay. However, compared to the scheduling delay, the propagation delay or packet processing delay is negligible, so they are not considered in this thesis.

Clearly, the transmit delay is largely affected by the scheduling policy implemented at the transmitter side. Hence, for evaluation of the system performance, especially for the RT system which is sensitive to latency, three metrics are mainly used in recent scheduling algorithms, and they are the instantaneous delay, the average delay and the delay variation. For example, in [9] the authors design the priority-based scheduling algorithm based on the instantaneous delay, and the delay variation is also investigated for evaluation of the algorithms. However, in [7] and [10], the proposed utility-based approaches consider the average delay metric, and focus on the average delay distribution

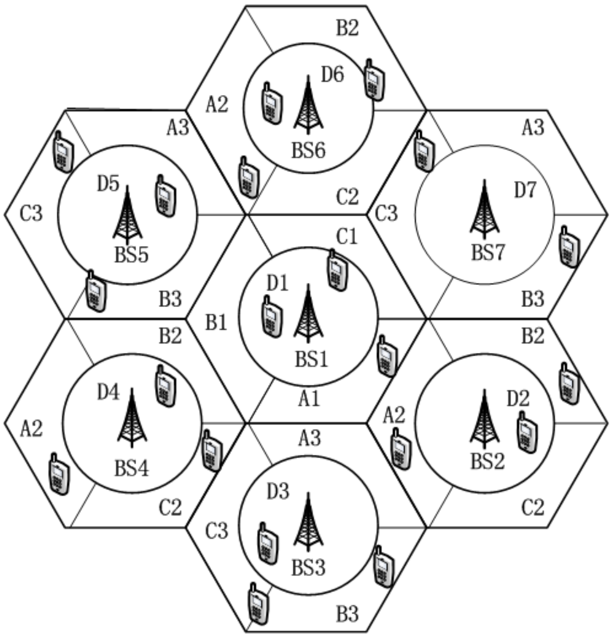


Figure 1.4: Hexagonal multi-cell networks.

[7] or delay violation distribution [10].

1.2.5 Coordinated Multipoint Networks

Hexagonal Multi-cell Cellular Networks

For a multi-cell network, a hexagonal multi-cell cellular network is often considered, see Fig.1.4. In the cellular network, a cell consists of a base station (BS), named the serving BS or anchor BS, and a set of randomly located mobile stations (MSs). Each cell could be further divided into three directional areas, named base station sectors (BSSs), and each BSS has one directional antenna. Based on the long term channel gain, MSs are classified as either in the cell-center or the cell-edge area. For instance, in Fig.1.4, Ai, Bi, Ci denote the three BSSs in the cell, respectively. Di refers to the cell-center area, while the area outside the circle boundary is the cell-edge area.

In conventional networks, MSs communicate only with the serving BS. ICI is incurred when the adjacent BSs transmit signals using the same spectrum. With the continuously increasing demands of emerging wireless applications, but scarce radio resources (i.e., time slot, bandwidth spectrum, space, and power), full frequency reuse and increasing BSs density will achieve higher spectral efficiency, but at the same time brings more sever ICI. In fact, when users are at the cell-edge areas, a full frequency reuse causes serious ICI, leading to sharply degradation of system throughput and fairness. Therefore, ICI becomes a major limiting factor for further improvement of system performance in the wireless networks [32].

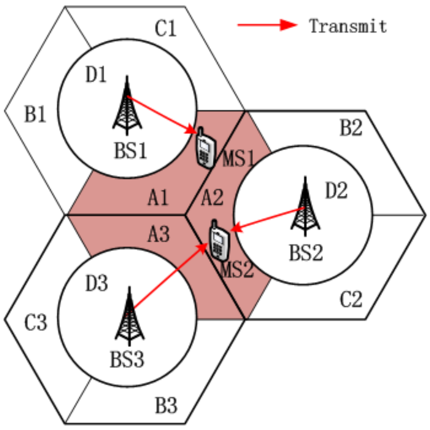
CoMP Transmission Schemes

Recently, CoMP transmission, also known as network coordination or base station coordination (BSC), has been identified as one of the key techniques for mitigating ICI, to improve the cell-edge performance. CoMP networks allow a group of BSs to simultaneously transmit/receive signals to/from a group of MSs, with each having an anchor BS or serving BS from this group of BSs. In downlink CoMP transmission, two transmission schemes are mainly considered: joint processing (JP) and coordinated scheduling/beamforming (CS/CB) [33].

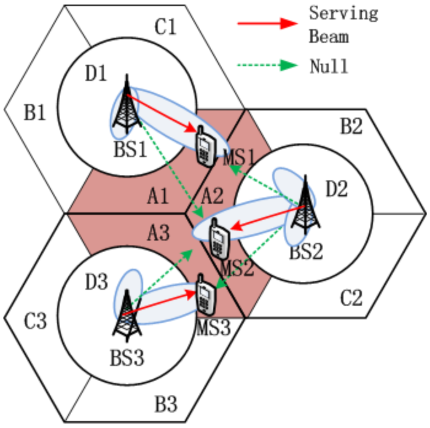
- **Joint Processing (JP)** is further categorized into joint transmission (JT) and dynamic cell selection (DCS). In JT, the signals are transmitted on the same spectrum simultaneously from multiple BSs among coordinated BSs (i.e., from non-serving BS as well as the serving BS). In principle, with perfect CSI, a subset of cells are selected for inter-cell coordination, in addition to individual selection so as to maximize the SINR at an MS set. Practically, JT is implemented based on codebook-based precoding, so as to reduce feedback signal overhead. Fig.1.5a shows the operation principle of JT in the downlink. MS2 is allocated at the cell-edge areas, having anchor BS2, and two neighboring BS1 and BS3. By JT, MS2 receives signals from both BS2 and BS3 simultaneously. Clearly, CSIs as well as user data symbols are shared among multiple coordinated BSs for implementation of JT.

- **Coordinated Scheduling/Beamforming** Unlike JP, with CS/CB MS is served only by its anchor BS. However, scheduling/beamforming is coordinated among multiple coordinated cells, so that ICI from neighboring BSs is reduced to the minimum level. As in Fig.1.5c, transmit beamforming weights for each MS are coordinated to mitigate the interference to other scheduled MSs within the coordinated BSs. Hence, the cell edge user throughput is improved due to the increase in received SINR. Similarly as in DSC, only channel states are required to be shared within the coordinated BSs.

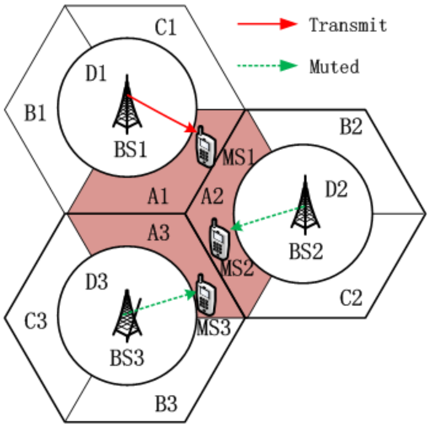
  In DSC, the BS with the minimum transmission pathloss among the coordinated BSs is dynamically selected, and then the other cells are muted (i.e., they do not transmit signals over the same spectrum), so the cell-edge user does not receive other interference from neighboring cells. Hence, the received signal power is maximized, and the interference from adjacent cells is mitigated. The operating principle of DCS is shown in Fig.1.5b. Since BS1 is serving MS1, the neighboring BS2 and BS3 are set to mute over the frequency resource so as to mitigate the ICI. Hence, DSC reduces backhauling overhead by only sharing channel states, but not the transmit data.



(a) JT.



(b) CS/CB.



(c) DSC.

Figure 1.5: CoMP downlink transmission schemes. a) JT. b) DCS. c) CS/CB.

Clearly, RRM cooperation among multiple cells plays an important role for mitigating ICI, and hence in turn enhancing the system performance in CoMP networks. Currently, RRM strategies are studied a lot for increasing average cell-edge throughput and balancing system efficiency and fairness [2], [4]-[6]. [2] proposes to select the set of cochannel users and the power allocation across tones to maximize the weighted system sum rate subject to per BS power constraint, assuming that only channel quality is shared among the coordinated APs. Hence, the scheduling and power allocation scheme allows only anchor BSs to transmit signals to the corresponding MSs, and SINR as well as the system performance is improved by ICI coordination among coordinated BSs. In [4], the cross-layer optimization problem of inter-cell coordinated resource allocation for OFDM system is addressed, and a heuristic algorithm to distribute radio resources among multi-users according to their individual QoS requirements is proposed to achieve better fairness. By exploiting inter-cell adaptive power allocation and ICI coordination, the algorithm is shown to provide enhancement in the throughput performance. A utility-based resource allocation algorithm with universal frequency reuse is proposed in [5] to support BSC transmission. For practical use, the entire network is divided into a number of disjoint coordinated clusters of BSSs in the algorithm to reduce backhauling and signaling overhead. In [6], a cooperative frequency reuse (CFR) scheme supporting joint transmission is proposed, and is proved to significantly improve the cell-edge throughput. In CFR transmission scheme, JT is fully used over a fraction of spectrum, so that the cochannel ICI is significantly mitigated and hence in turn the channel SINR is maximized.

1.3 System Model

In this section, we introduce the system model used in our thesis. Firstly, we describe the static clustered CoMP network, and the expression of the instantaneous SINR is given. Then, the bit allocation scheme based on MCS adaptation strategy is presented, as well as the data rate expression. Thirdly, the traffic models for different services are described, and a two-state Markov process is used to model the VoIP traffic considered in our thesis. Then a description is presented for the structure of our downlink packet scheduling system, as well as the queue structure assumed in this thesis. Finally, ant colony system is introduced for addressing combinatorial optimization problem, and a traditional ACO graph model is described.

1.3.1 Clustered CoMP Networks

In our thesis, we consider a hexagonal multi-cell cellular network as in Fig.1.4. In order to reduce signaling and backhauling overhead user grouping is used, serving only subsets of terminals with a subset of BSs. We divide the network into smaller subsystems [5]. In this thesis, all the available BSSs in the network are separated into a number of disjoint clusters of coordinated BSSs. Each cluster consists of three adjacent BSSs. See Fig.1.1 for an example of the CoMP network. As in Fig.1.6, BSSs $A1, A2$, and $A3$ are grouped into a CoMP cluster. We then focus on the downlink of the clustered CoMP system. A central unit (CU) is assumed to be employed in each cluster for user scheduling and power control for the coordinated set of BSSs attached to the cluster. MSs within each cell are divided into two classes according to their long term pathloss, classified as cell-center MSs and cell-edge MSs. In our thesis, only cell-edge MSs (shortly called MSs) are taken into account. Each BSS is configured with one directional antenna with maximum transmit power P, and each MS is equipped with one receive antenna. Assume the bandwidth is fully reused among all the BSSs. In addition, we consider non-coherent CoMP transmission, i.e., the received signals for a CoMP user are non-coherently added up at the receiver side.

Assume there are a set of BSSs $\mathcal{N}$ and a set of MSs $\mathcal{M}$ in the network. Let $\mathcal{M}_n$ denote the MS set in the BSS n. Hence, $\mathcal{M}_n$ is a subset of $\mathcal{M}$. Given the cardinality

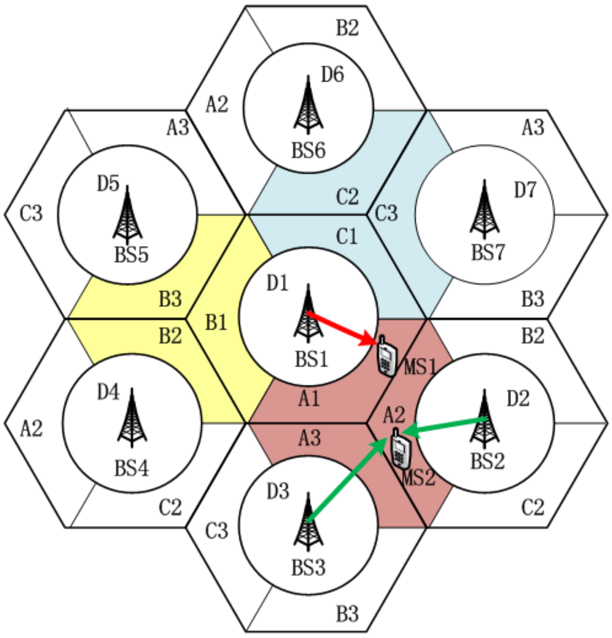


Figure 1.6: CoMP clusters.

of the BSS set as $|\mathcal{N}| = N$, and that of the subset $\mathcal{M}_n$ as $|\mathcal{M}_n| = M_n$, thus the entire network has a total number of users of $M = \sum_{n=1}^{N} M_n$. K subchannels are available for resource allocation, and frequency reuse factor is one. For simplicity, we assume that each subchannel of a BSS can only be assigned to no more than one MS, and the power assigned to each subchannel is fixed.

In conventional cellular networks where each MS exchanges data only with the serving BSS, hence, the discrete-time baseband signal received by MS m of its serving sector BSS n on the k^{th} subchannel is

$$r_m^k = \underbrace{\sqrt{P_n^k} H_{n,m}^k s_{n,m}^k}_{\text{signal of interest}} + \underbrace{\sum_{j \neq n} \left(\sqrt{P_j^k} H_{j,m}^k \left(\sum_{s \in \mathcal{M}, s \neq m} s_{j,s}^k \right) \right)}_{\text{inter-cell interference}} + \underbrace{w_m^k}_{noise}, \tag{1.9}$$

where P_n^k is the transmit power of BSS n on the k^{th} subchannel. $s_{n,m}^k$ is the complex symbol transmitted from BSS n to MS m over the subchannel k. $H_{n,m}^k$ is the complex channel gain between BSS n and MS m, consisting of path loss, shadow fading and fast fading. w_m^k is the white Gaussian noise with noise power N_0. Hence, the instantaneous SINR on the k^{th} subchannel is

$$\gamma_m^k = \frac{P_n^k G_{n,m}^k}{\sum_{j \neq n} P_j^k G_{j,m}^k + N_0}, \tag{1.10}$$

where $G_{n,m}^k = |H_{n,m}^k|^2$. By contrast, in a CoMP network supporting non-coherent joint transmission, each MS can receive signals from any combination of the three BSSs in a cluster. Define the set of cooperative BSSs in the cluster as coordinated BSS set (CBS), then the CBS of MS m is denoted by D_m. $\bar{D}_m$ is the complement set of D_m, and thus is the set of ICI sources of MS m. Hence, for the CoMP network, the instantaneous received SINR on the k^{th} subchannel for MS m is defined as

$$\gamma_m^k = \frac{\sum_{n \in D_m} P_n^k G_{n,m}^k}{\sum_{j \in \bar{D}_m} P_j^k G_{j,m}^k + N_0}. \tag{1.11}$$

1.3.2 Bit Allocation Scheme

We consider rate adaptation strategy for throughput maximization based on block-error-rate (BLER) requirements instead of BER requirements. The 16-level MCSs are listed in the Table. 1.1, and the block-error-rate (BLER) curves for all the 15 CQIs are plotted in Fig.1.7. Hence, provided the instantaneous CSI feedback, and in turn the instantaneous SINR γ, according to (1.2), the throughput expectation with the optimal MCS adaptation is given by

$$\bar{\eta}(\gamma) = R^{(l^*)} \left(1 - \mathrm{BLER}^{(l^*)}(\gamma)\right), \tag{1.12}$$

where l^* is the optimal MCS obtained by

$$l^* = \arg\max_l \left\{ R^{(l)}(1 - \mathrm{BLER}^{(l)}(\gamma)) \right\}. \tag{1.13}$$

Assume the target BLER is BLER_Q, and in turn the required SNR ${\gamma_Q}^{(l)}$ is obtained by the inverse function of (1.3). The MCS schemes are ordered such that $R^{(l-1)} < R^{(l)} < R^{(l+1)}$, so an MCS is optimal if it is in the SNR range [14]

$$\gamma \in \left[{\gamma_Q}^{(l)}, {\gamma_Q}^{(l+1)}\right). \tag{1.14}$$

CQI value	Modulation scheme	Channel coding rate	CQI value	Modulation scheme	Channel coding rate
0	N/A	N/A	8	16QAM	490/1024
1	QPSK	78/1024	9	16QAM	616/1024
2	QPSK	120/1024	10	64QAM	466/1024
3	QPSK	193/1024	11	64QAM	567/1024
4	QPSK	308/1024	12	64QAM	666/1024
5	QPSK	449/1024	13	64QAM	772/1024
6	QPSK	602/1024	14	64QAM	873/1024
7	16QAM	378/1024	15	64QAM	948/1024

Table 1.1: 16-level MCSs.

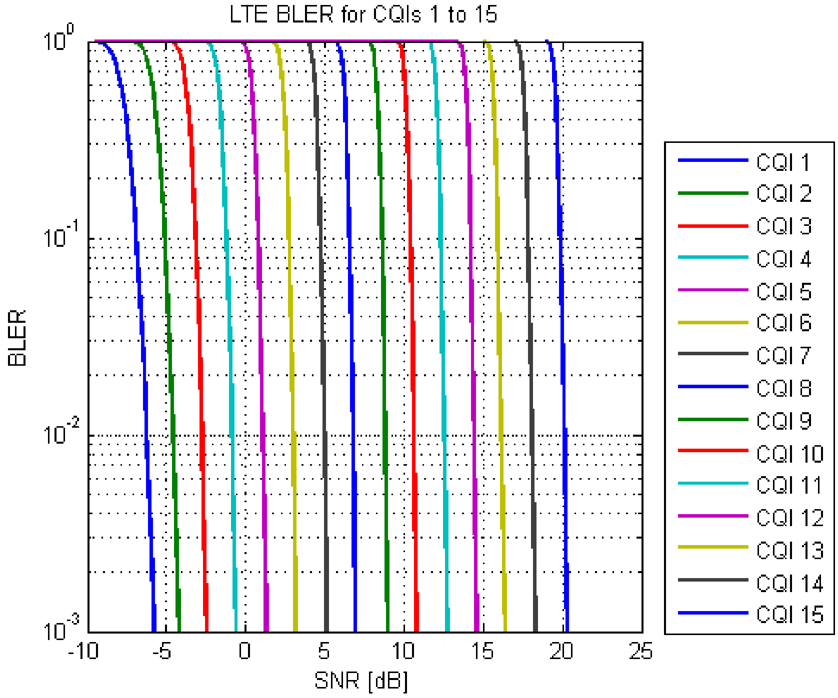


Figure 1.7: BLER curves for all 15 CQI values.

(1.13) and (1.14) imply that MCS adaptation is to select the optimal MCS subject to the target BLER requirement. Analytically, the envelope curve of all possible throughputs with different MCS schemes can be approximated by a modified formula of (1.1) [14]

$$\bar{\eta}(\gamma) = B \log_2(1 + \beta\gamma), \tag{1.15}$$

where β is calculated based on equivalent bit error rate (BER)

$$\beta = \frac{-1.5}{\ln(5\mathrm{BER})}. \tag{1.16}$$

For simplicity, we use the expression (1.15) to approximate the throughput, so the transmit rate of MS m on the subchannel k of the CoMP system model introduced in previous subsection is given by

$$R_m^k = B \log_2 (1 + \beta\gamma_m^k), \tag{1.17}$$

where γ_m^k is the SINR of MS m on the subchannel k, and B is the system bandwidth. Hence, the instantaneous data rate for MS m becomes

$$R_m = \sum_{k=1}^{K} R_m^k. \tag{1.18}$$

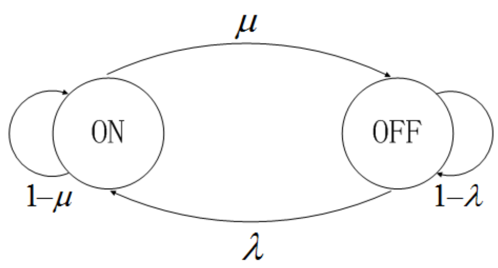


Figure 1.8: Two state Markov chain for VoIP period transitions.

1.3.3 Traffic Models

In our thesis, we consider two different types of users in our system, i.e., best effort users and real time users. A full buffer scenario, where there is always data available at the base station for all terminals in the cell, is assumed for each BE user. For RT users, we consider VoIP traffic model with bursty low bit rate. A two state Markov process is used to model VoIP as illustrated in Fig.1.8. The probability of state transition from activity (*ON* periods) to silence (*OFF* periods) and from silence to activity are μ and λ respectively. During the *ON* period, we assume that fixed sized packets are generated at a constant interval of T_{on}, and the constant packet rate is assumed as α_{on}. Relatively during the *OFF* period, no data is generated, and the packet interval is set as T_{off}. The average bit rate associated with the *ON-OFF* process as in Fig.1.8 is then given by

$$\bar{\alpha} = \frac{T_{on}}{T_{on} + T_{off}} \alpha_{on}. \tag{1.19}$$

A maximum tolerable instantaneous packet delay for VoIP is imposed as $d_{\max}$. A packet exceeding the maximum allowable delay is discarded. Additionally, packets experiencing poor channel and unsuccessfully received are dropped at the receiver side. We define packet loss as aggregate of both time out packets and dropped packets. Acceptable VoIP service is maintained for a packet loss ratio (PLR) less than 2%. A user is supposed in QoS outage if it exceeds this ratio.

According to the diverse characteristics of the two different traffic models, we classify the users in our system into two categories, i.e., BE users and VoIP users, and the sets of these two types of users are denoted by $\mathcal{M}_{BE}$ and $\mathcal{M}_{BE}$ respectively, where $\mathcal{M}_{BE} \cup \mathcal{M}_{BE} = \mathcal{M}$. For simplicity, we assume that each user can only receive one specified service during the whole simulation period, namely, $\mathcal{M}_{BE} \cap \mathcal{M}_{BE} = \varnothing$.

1.3.4 Packet Scheduling System

The downlink packet scheduling system considered in this thesis consists of three blocks, including a packet classifier (PC), a buffer management block (BMB), and a packet scheduler (PS) as shown in Fig.1.9 [34]. Assume each user has an individual buffer at the BS.

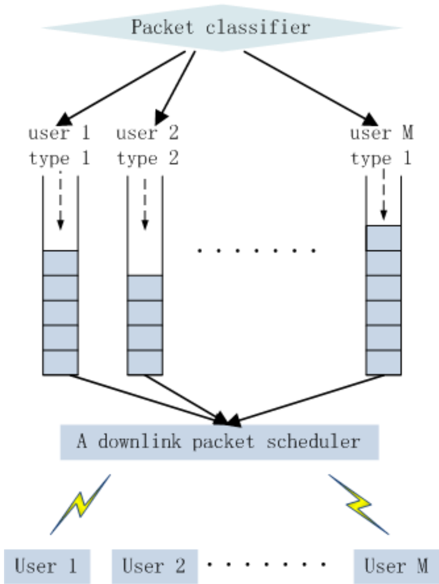


Figure 1.9: Structure of the downlink packet scheduling system.

The PC classifies incoming packets according to their traffic types and forwards them to corresponding receiving buffers in BMB. The BMB maintains QoS statistics such as the arrival time and the delay deadline of each packet, the average arrival rate, the packet numbers, the queue size, and the head-of-line (HOL) delay in each buffer. Additionally, the BMB drops packets that exceed the maximum allowable delay, and control the segmentation of packets in the buffers. Finally, the PS calculates the priority for each users and transmits packets to users according to the scheduling priority obtained based on channel status feedbacks and QoS statistics maintained in BMB, and makes a joint scheduling and subchannel assignment once every slot.

Clearly, the PS that operates at the MAC layer is the key component for delivering services to users in the system. For a wireless network with a variety of services and applications with various QoS requirements, many metrics can be used to evaluate a system. In our thesis, we consider two different utility functions for the two different types of users, i.e., BE and VoIP users. For BE users, the utility is usually defined with respect to data rate. Hence, we define the utility function as a concave and monotonically increasing function of average data rate, namely

$$U_m(t) = U_{BE,m}\left(\bar{R}_{BE,m}(t)\right), \tag{1.20}$$

where $\bar{R}_{BE,m}(t)$ is the average data rate of BE user m at time slot t, and can be estimated using an exponential filter [35]

$$\bar{R}_{BE,m}(t) = (1-\rho_B)\bar{R}_{BE,m}(t-1) + \rho_B R_{BE,m}(t). \tag{1.21}$$

$0 < \rho_{_B} < 1$ is a constant given by $\rho_{_{BE}} = T_s/T_{BE}$, where T_{BE} is the filter window size. We assume that VoIP user m is associated with its average queuing delay $\bar{d}_{V,m}(t)$, and the corresponding utility function is [31]

$$U_m(t) = U_{V,m}\left(\bar{d}_{V,m}(t)\right). \tag{1.22}$$

Obviously, with longer delay, the user has a lower level of satisfaction. It is reasonable to assume that $U_{V,m}\left(\bar{d}_{V,m}(t)\right)$ is monotonically decreasing and concave. Let $Q_{V,m}(t)$ be the amount of bits in the queue of VoIP user m at time t. Suppose serving rate at time t for VoIP user m is $R_{V,m}(t)$, and arrival packet at time t has the size of $\alpha_{V,m}(t)$, so the queue size of VoIP user m at time slot t is expressed as

$$Q_{V,m}(t) = Q_{V,m}(t) - \min\left\{Q_{V,m}(t-1), R_{V,m}(t)T_s\right\} + \alpha_{V,m}(t-1). \tag{1.23}$$

Similarly as in (1.21), assume a time window with length T_V, and then the average queue length over the time window of user m at time t is given by

$$\bar{Q}_{V,m}(t) = (1-\rho_V)\bar{Q}_{V,m}(t-1) + \rho_V Q_{V,m}(t), \tag{1.24}$$

where $\rho_V = T_s/T_V$, is a constant in (0,1). Hence, the average queuing delay $\bar{d}_{V,m}(t)$ at time t can be defined as

$$\bar{d}_{V,m}(t) = \frac{\bar{Q}_{V,m}(t)}{\bar{\alpha}_{V,m}}, \tag{1.25}$$

where $\bar{\alpha}_{V,m}$ denotes the time averaging arrival bits of VoIP user m, and can be estimated by (1.19).

1.3.5 Ant Colony Optimization

ACO has been first proposed by Marco Dorigo in 1990s to address combinatorial optimization problems [13][14]. In the ACO, a set of cooperating agents called ants cooperate to find good solutions, mimicking the behavior of real ants in search for food. Ants cooperate by leaving some amount of pheromone on the path to the food. At each decision point, each ant chooses randomly a path to go on. Different edges have different probability of being selected. Since ants move at approximately a constant speed, the ants passing by shorter path reach the food faster than those that choose the longer ones. Hence, pheromone accumulates at a higher rate on the shorter path, which makes it more attractive. In Fig.1.10, we show a way ants exploit pheromone. By using this idea a meta-heuristic method was proposed, named ant colony optimization [13], to find a shortest path between two points.

In an ACO system, all the possible paths from the nest to the food are made from a number of vertexes and edges. The ants choose the next edge in each vertex. The solution of the optimization problem is a list of edges and vertexes which make a shortest trail from the nest to the food. In Fig.1.11, we show a simple case of an ACO graph. Ants choose their path based on two factors: *visibility* and *trail intensity*. Visibility (V) is

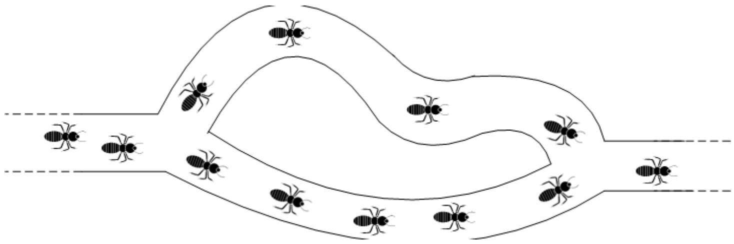


Figure 1.10: An example of ACO system.

what an ant can see, and usually it is considered as the inverse of the edge length. Trail intensity (T) is the pheromone left at each vertex, which is related to the number of ants passing by the vertex [15]. An ant selects its own path randomly. According to the state transition rule, named random-proportional rule, the probability for an ant k in vertex i to choose the edge to the vertex j is

$$p_{i,j}^{k} = \frac{[T_{i,j}^{k}]^{\alpha} \cdot [V_{i,j}^{k}]^{\beta}}{\sum_{m=1}^{M} [T_{i,j}^{k}]^{\alpha} \cdot [V_{i,j}^{k}]^{\beta}}, \tag{1.26}$$

where $T_{i,j}^{k}$ and $V_{i,j}^{k}$ are the trail intensity and the visibility of the edge between vertex i and vertex j, respectively, when ant k passes by. M is the set of vertexes, and here in this case, it is a set of all the vertexes in Fig.1.11. α and β are two constant. α is usually set to 1, and β is a parameter which determines the relative importance of pheromone versus distance.

Once all ants have completed their tours a global pheromone updating rule is applied [14]. A fraction of the pheromone evaporates on all edges. Hence, on the edges which are part of many short trails, more pheromone would be deposited by ants, and would become more attractive. The process is iterated, until all the ants follow the shortest path from the nest to the food.

1.4 Purpose and Contribution

The research problem considered in this thesis is how to effectively assign subcarriers and allocate power on the downlink of the CoMP network by exploiting knowledge of the wireless channel conditions and the characteristics of traffic to improve the spectral efficiency and guarantee diverse QoS. This section describes the contributions and the idea behind the contributions of this thesis. The method used in this thesis is system level simulation. The CoMP LTE simulator used in this thesis is implemented based on the LTE system level simulator [36][37]. Our proposed algorithms are simulated and evaluated on the system level, and all the results regarding the performance of the proposed joint scheduling schemes are produced by system level simulation.

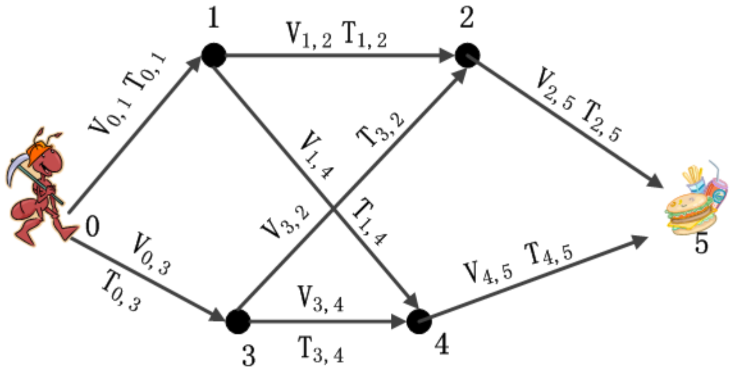


Figure 1.11: A simple case of an ACO graph.

1.4.1 Utility-Based Joint Scheduling and Power Control

As we want to increase the performance of mixed-traffic CoMP networks, utility functions for different traffic pattern are modeled to present the satisfaction level of different types of users. We simplify the system model, and focus on flat fading channel in the downlink of single-cluster CoMP system, which indicate that no inter-cluster interference is considered. An optimal utility-based joint packet scheduling algorithm is proposed to maximize the sum utility in the CoMP network. Through analytical derivation, we reformulate the non-linear optimization problem into a feasible linear optimization problem. Besides, binary power control (i.e., in any time slot, the cell either transmits with full power or does not transmit) is also employed for power saving.

To evaluate the performance of the proposed algorithm, we also investigate the proportional-fair scheduling without joint transmission, and the proportional-fair based joint transmission scheme. Additionally, we also look into the performance of utility-based scheduling without BSS cooperation. We evaluate the cell-edge sum utility, average throughput of BE users, and the VoIP call outage ratio, i.e., the interruption probability of VoIP call due to excessive packet delay. By system level simulation, we show that compared with the three algorithms, our proposed algorithm substantially improve the system performance at the cell-edge area, by providing better system throughput of BE users, and meanwhile suppressing the call outage probability for VoIP users by taking advantage of joint transmission and diverse utility functions.

1.4.2 ACO-Based Two-Step Joint Scheduling

As the extension to the work based on the utility-based joint scheduling algorithm proposed for flat fading single-cluster network, we also propose an ACO-based two-step joint scheduling algorithm for multi-cluster OFDMA CoMP systems. In this contribution, we take into account more factors, i.e., inter-cluster interference, and multiple subchannels, and we also investigate cases with heavy traffic load, i.e., when the number of cell-edge users is large. To solve this combinatorial optimization problem is usually computationally prohibitive. Hence, we introduce the ant colony algorithm to address this optimization problem. On each subchannel, the combinatorial user selection problem is analytically modeled into a pseugraph with each vertex representing a BSS in the cluster based on utility. Then the subchannels are assigned to each group of coordinated BSSs using a greedy algorithm. The computational complexity for the proposed algorithm is also discussed in one of the included paper.

The ACO based joint scheduling algorithm is compared with the optimal algorithm and a greedy user selection based algorithm by system level simulation in our thesis. The average throughput, and packet drop ratio are investigated for BE users and VoIP users, respectively. The cell-edge sum utility is also measured. Besides, we also calculate the ratios of the numbers of QoS-satisfied users to the total number of cell-edge users for both BE and VoIP users, respectively. Via simulation results, we show that our proposed algorithm well approximates the optimal solution, and yield a better performance compared to the greedy algorithm.

1.5 Included Papers

Paper A - A Utility-Based Scheduling Approach for Multiple Services in Coordinated Multi-Point Networks

The main purpose of this paper is to address the scheduling problem in multi-cell multi-service networks. A utility-based coordinated transmit scheduling method is proposed to support joint transmission in CoMP systems with mixed VoIP and BE traffic patterns. The objective is to improve the cell-edge performance for VoIP users as well as BE users. The optimization problem is analytically formulated, which amounts to a user-group selection problem, i.e., choosing the best user group for each time slot. Via simulation results, we show that compared to traditional scheduling schemes, the proposed algorithm significantly improve the throughput of cell-edge BE users, while suppressing the call outage ratio.

Paper B - Joint Scheduling for Multi-Service in Coordinated Multi-Point OFDMA Networks

In this paper, the issues upon user scheduling in CoMP OFDMA networks are addressed. We focus also on the downlink packet transmission with multiple services, i.e., BE and VoIP services. In addition, we also take into account of multiple clusters, inter-cluster interference, as well as multiple subchannels. To improve cell-edge performances and guarantee diverse quality of service, a utility-based joint scheduling algorithm is proposed.

The proposed algorithm consists of two steps: ant colony optimization based joint user selection and greedy subchannel assignment. Compared with the optimal algorithm and the greedy user selection based algorithm, we show via simulation results that our proposed algorithm well approaches the optimal solution, and yield better performance than the greedy algorithm by taking advantage of ant colony optimization on each subchannel.

1.6 Future Work

Some open issues related to the research work presented in this thesis are listed in this section. The work presented in this thesis is based on a number of simplifying assumptions. The model of the wireless system is supposed to include more realistic assumption for future work, which, for example, should consider issues as:

- a more precise model of the signaling overhead for communication of coordinated base stations.
- inperfect CQI at the BS side.
- transmission errors in the feed-back channels.
- a more precise model of VoIP traffic than two-state Markov model.
- practical implementation limitations.

All items above will most likely decrease performance of the scheduling algorithms, and are worth being taken into account in future work.

Another topic of interest would be to investigate and apply ant colony optimization to not only joint user selection on each subchannel, but also to subchannel assignment, i.e., to model the ant system with each vertex representing a combination of user, subchannel and coordinated base station sector. In addition, coherent joint transmission as well as coordinated scheduling could be also investigated and proposed for multi-service CoMP systems.

References

[1] 3GPP TR 36.814, V9.0.0, Further advancements for EUTRA physical Layer Aspects, Mar. 2010.

[2] L. Venturino, N. Prasad and X. Wang, Coordinated scheduling and power allocation in downlink multicell OFDMA networks, *IEEE Trans. on Veh. Technol.*, vol. 58, no. 6, pp. 2835-2848, Jul. 2009.

[3] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, Nortel, Interference coordination and cancellation for 4G networks, *IEEE Comm. Mag.*, vol. 47, no. 4, pp. 74-81, Apr. 2009

[4] L. Chen, W. Chen, X. Zhang, Y. Zhang, C. Xiong and D. Yang, Intercell coordinated resource allocation for mobile WiMAX system, *Proc. IEEE WCNC*, pp. 1-6, Apr. 2009.

[5] J. Li, X. Xu, X. Chen, X. Tao, H. Zhang, T. Svensson, and C. Botella, Downlink radio resource allocation for coordinated cellular OFDMA networks, *IEICE Trans. Comm.*, vol. e93-b, no. 12, pp. 3480-3488, 2010.

[6] J. Li, H. Zhang, X. Xu, X. Tao, T. Svensson, C. Botella, B. Liu, A Novel Frequency Reuse Scheme for Coordinated Multi-Point Transmission, *IEEE Veh. Technol. Conf.*, pp. 1-5, May. 2010.

[7] X. Wang, G. B. Giannakis, and A. G. Marques, A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks, *Proc. IEEE*, vol. 95, no. 12, pp. 2410-2431, Dec. 2007.

[8] J. Puttonen, N. Kolehmainen, T. Henttonen, M. Moisio, and M. Rinne, Mixed traffic packet scheduling in UTRAN Long Term Evolution Downlink, *Proc. PIMRC'08*, pp.1-5.

[9] M. Katoozian, K. Navaie, and H. Yanikomeroglu, Utility-based adaptive radio resource allocation in OFDM wireless networks with traffic prioritization, *IEEE Trans. Wireless Comm.*, vol. 8, no. 1, 2009.

[10] G. Song, Y. Li, L. J. Cimini, and H. Zheng, Joint channel-aware and queue-aware data scheduling in multiple shared wireless channels, *IEEE WCNC*, vol. 3, pp. 1939-1944, Jul. 2004.

[11] H. Skjevling, D. Gesbert, and A. Hjrungnes, Low-complexity distributed multibase transmission and scheduling, *EURASIP J. Adv. Sig. Proc.*, 2008.

[12] R. Y. Chang, Z. Tao, J. Zhang, and C.-C. Jay Kuo, Multicell OFDMA downlink resource allocation using a graphic framework, *IEEE Trans. on Veh. Technol.*, vol. 58, Sep. 2009.

[13] M. Dorigo, T. Stutzle, Ant colony optimization, *MIT Press*, Cambridge, MA, 2004.

[14] M. Dorigo, and L. M. Gambardella, Ant colony system: a cooperative learning approach to the travelling salesman problem, *IEEE Trans. Evol. Comput.*, vol. 1, 1997.

[15] H. Ahmadi, Y. H. Chew, and C. C. Chai, Multicell Multiuser OFDMA dynamic resource allocation using ant colony optimization, *IEEE Veh. Technol. Conf.*, Jul. 2011.

[16] H. Ahmadi, Y. H. Chew, Subcarrier-and-bit allocation in multiclass multiuser single-cell OFDMA systems using ant colony optimization based evolutionary algorithm, *IEEE WCNC Proc.*, 2011.

[17] Y. Zhao, X. Xu, Z. Hao, X. Tao, and P. Zhang, Resource Allocation in multiuser OFDM System Based on Ant Colony Optimization, *IEEE WCNC*, Jul. 2010.

[18] G. Andrea, *Wireless Communications*, Cambridge University Press, 2005.

[19] A. J. Goldsmith, and S.-G. Chua, Adaptive coded modulation for fading channels, *IEEE Trans. Comm.*, vol. 46, no. 5, pp. 595-602, May. 1998.

[20] A. J. Goldsmith, and Chua, S. G., Variable-rate variable-power MQAM for fading

[21] S. Nanda, K. Balachandran, and S. Kumar, Adaptation techniques in wireless packet data services, *IEEE Comm. Mag.*, pp. 54-64, Jan. 2000.

[22] J.M. Holtzman, CDMA forward link waterfilling power control, *Proc. IEEE Veh. Technol.*, vol. 3, pp. 1663-1667, May. 2000.

[23] J.M. Holtzman, Asymptotic analysis of proportional fair algorithm, *Proc. PIMRC'01*, vol. 2, pp. 33-37, 2001.

[24] R. Jain, *The art of computer systems performance analysis*, Wiley, New York, 1991.

[25] 3GPP TR 25.814, V7.1.0, Physical layer aspects for evolved UTRA, Sep. 2006.

[26] I. Koutsopoulos, and L. Tassiulas, Volos, Cross-Layer Adaptive Techniques for Throughput Enhancement in Wireless OFDM-Based Networks, IEEE/ACM Trans. on Networking, vol. 14, pp. 1056-1066, Oct. 2006.

[27] S. Pietrzyk, and G.J.M. Janssen, multiuser subcarrier allocation for QoS provision in the OFDMA systems, *Proc. IEEE Veh. Technol.*, vol. 2, pp. 1077-1081, Dec. 2002.

[28] Y. Wei, and R. Lui, Dual methods for nonconvex spectrum optimization of multicarrier systems, *IEEE Trans. on Comm.*, vol. 54, pp. 1310-1322, Jul. 2006.

[29] Z. Jiang, Y. Ge, and Y. G. Li, Max-utility wireless resource management for best-effort traffic, *IEEE Trans. on Wireless Comm.*, vol. 4, pp. 100-111, Jan. 2005.

[30] S. Shenker, Fundamental design issues for the future Internet, *IEEE J. Select. Areas Comm.*, vol. 13, pp. 1176-1188, Sept. 1995.

[31] G. Song, *Cross-layer resource allocation and scheduling in wireless multicarrier networks*, Ph.D. dissertation, Georgia Inst. Technol., Apr. 2005.

[32] R.Y. Chang, Z. Tao, J. Zhang, and C.-C.J. Kuo, Multicell OFDMA Downlink Resource Allocation Using a Graphic Framework, *IEEE Trans. on Veh. Technol.*, vol. 58, pp. 3494-3507, Sep. 2009.

[33] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, Coordinated multipoint transmission/reception techniques for LTE-advanced, *IEEE Wireless Comm.*, vol. 17, pp. 26-34, Jun. 2010.

[34] A. I. Pérez-Neira, and M. R. Campalans, *Cross-Layer Resource Allocation in Wireless Communications : Techniques and Models from PHT and MAC Layer Interaction*, Academic Press, 2009.

[35] M. Nicolaou, A. Doufexi, S. Armour, and Y. Sun, Scheduling Techniques for Improving Call Capacity for VoIP Traffic in MIMO-OFDMA Networks, *IEEE Veh. Technol.*, Sep. 2009.

[36] Institute of Communications and Radio-Frequency Engineering, Vienna LTE Simulators System Level Simulator Documentation, v1.3r427.

[37] J. C. Ikuno, M. Wrulich, and M. Rupp, “System level simulation of LTE networks,” in Proc. 2010 IEEE 71st Vehicular Technology Conference, Taipei, Taiwan, May 2010.

PAPER A

A Utility-Based Scheduling Approach for Multiple Services in Coordinated Multi-Point Networks

Binru Huang, Jingya Li, Tommy Svensson

Published in
14th International Symposium on
Wireless Personal Multimedia Communications (WPMC)

Brest, France, 2011

Abstract

Coordinated Multi-Point (CoMP) joint transmission is considered as a key technique to mitigate inter-cell interference (ICI) and improve the cell-edge performance in 3GPP LTE-Advanced. In this paper, a utility-based coordinated transmit scheduling method is proposed to support joint transmission in CoMP systems. The objective is to improve the cell-edge performance in downlink packet-based CoMP networks with mixed real-time voice over IP (VoIP) and best-effort (BE) traffic patterns. Via simulation results we show that compared to the traditional scheduling schemes with neither CoMP joint transmission nor diverse quality of service (QoS) provisioning, the proposed algorithm in this paper improves the cell-edge efficiency of BE users by greater than 45%, while better satisfying the QoS requirements for VoIP users with significant decrease in call outage by greater than 98%.

Keywords: Coordinated multi-point (CoMP), OFDMA, multiple services, quality of service (QoS), scheduling and resource allocation.

2.1 Introduction

Coordinated Multi-Point (CoMP) joint transmission, also known as network coordination, is proposed as a promising technique to satisfy the system requirements regarding coverage, cell-edge user throughput and system efficiency [1]. In CoMP systems, multiple coordinated cells are connected via a high-speed backbone. By using joint transmission scheme in the downlink of CoMP systems, the inter-cell interference (ICI) can be significantly mitigated by applying the signals transmitted from other cells to assist the transmission instead of acting as interference.

Clearly, radio resource management (RRM) cooperation among multiple cells plays a key role for controlling ICI, and in turn improving the system performance in CoMP networks. Currently, RRM schemes for CoMP systems are mainly studied for increasing cell-edge throughput and achieving the balance between efficiency and fairness, assuming all the users in the network have the same traffic modes [2]-[4]. In [2] and [3], algorithms utilizing coordinated scheduling and power control are proposed for controlling ICI. In [4], a utility-based algorithm for multi-cell coordinated resource allocation is proposed, and is proved to be more efficient in improving the cell-edge average throughput and user fairness, compared to traditional single-cell transmission. Note that all these schemes treat the users equally without considering multiple traffic types in the CoMP networks.

However, beyond efficiency and robustness to ICI, present and next-generation wireless networks are challenged to meet the diverse QoS imposed by various services [5]. The research on mixed-traffic scenarios is receiving more attention due to its significance in practical deployment of the Evolved UTRAN (E-UTRAN). Currently, research focuses mainly on single-cell mixed-traffic scenarios [5]-[7]. In [5], a unified approach based on utility functions to QoS-guaranteed scheduling is proposed for time-division multiplexing (TDM) in downlink. In [6] a mixed best-effort (BE) and voice over IP (VoIP) traffic is studied, and a dynamic packet scheduling architecture is proposed to differentiate scheduling of different traffic classes. The result in [6] shows that with VoIP prioritizing keeps the VoIP UEs satisfied at the cost of decreased system spectral efficiency. In [7], a utility-based optimization is proposed, and is shown to be able to satisfy the delay requirement of real-time (RT) traffic while balancing the fairness and efficiency. The main limitation of [5]-[7] is that RRM is designed for the single cell scenario and no joint transmission is undertaken.

In this paper, we propose a joint packet scheduling and power control algorithm on flat fading channel for CoMP networks with mixed BE and VoIP services. We focus on the downlink of a CoMP cluster, consisting of three base station sectors (BSSs) with fixed maximum transmit power. The objective is to maximize the sum of all the users' utilities, which present the users' satisfaction levels based on different traffic. Binary power control (i.e., in any time slot, the cell either transmits with full power or does not transmit) is also used. The results show that the optimization problem amounts to a user-group selection problem, i.e. choosing the best user group for each time slot. Via the simulation results, we show that through diverse utility functions the proposed algorithm well satisfies users of different traffic types regarding QoS requirements. By taking advantage of the joint transmission scheme, the presented algorithm is also proved to significantly improve the cell-edge BE users' efficiency, while suppressing the call outage probability (i.e. the interruption probability of VoIP calls due to excessive packet delay) for VoIP service.

The rest of this paper is organized as follows. In section 2.2, we provide the system model considered in this paper. In section III, the constrained optimization objective is formulated, and a radio resource allocation algorithm for utility-based joint scheduling and power allocation is proposed. Simulation results are presented in section IV, and conclusions are presented in section V.

2.2 System Model

We focus on the downlink of a static CoMP cluster consisting of three neighboring BSSs. A central unit (CU) is used to determine user schedule and power control for all BSSs in the cluster; see Fig.1. According to the long term channel gain, users are divided into two classes, namely cell-center users (CCUs) and cell-edge users (CEUs). Joint transmission can only be applied to CEUs. In this paper, we focus only on CEUs. The BSSs are assumed to have one directional transmit antenna each with the same fixed maximum transmission power P, and share the same cell-edge bandwidth B. The CEUs are further divided into two categories based on the services they require, i.e. BE users and VoIP users, denoted as $\mathrm{CEU_{BE}}$ and $\mathrm{CEU_{V}}$ respectively in Fig.1. Each CEU is equipped with one receive antenna and can receive signals from a subset of the BSSs of the CoMP cluster.

The CoMP cluster is supposed to have a set $\mathcal{M}$ of CEUs and a set $\mathcal{N}$ of BSSs. In each time slot, the CU allocates users for each BSS n based on the channel state information (CSI) of each CEU m. A user schedule index $x_{nm}(t)$ is defined as

$$x_{nm}(t) = \begin{cases} 1, & \text{if BSS } n \text{ transmits to CEU } m \text{ at time slot } t; \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Hence, the user schedule matrix can be denoted as $\mathbf{X}(t) = [x_{nm}(t)]$ with size $N \times M$, where N and M are the cardinality of $\mathcal{N}$ and $\mathcal{M}$ respectively. Assume that each BSS can transmit to no more than one user in a time slot, and thus we have $\sum_{m=1}^{M} x_{nm}(t) \leq 1; \forall n \in \mathcal{N}$.

Let $P_n(t)$ denote the transmit power of BSS n at time slot t. Based on the binary power control assumption, each BSS either transmits with maximum power, i.e., P or does not transmit. Let $G_{nm}(t)$ denote the channel gain between BSS n and CEU m at time slot t, consisting of path-loss, shadow fading, and small-scale fading. Then with the power of the additive white Gaussian noise (AWGN) N_0, the signal to interference and

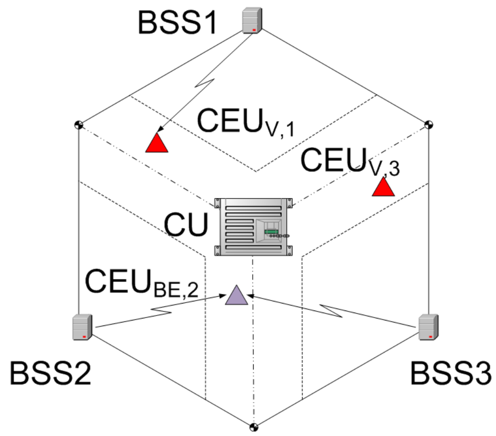


Figure 2.1: System model for downlink joint transmission in one CoMP cluster.

noise ratio (SINR) of the CEU m at time t based on non-coherent reception becomes

$$\gamma_m(t) = \frac{\sum_{n=1}^{N} G_{nm}(t) P x_{nm}(t)}{\sum_{j=1}^{N} G_{jm}(t) \left(\sum_{i \in \mathcal{M}, i \neq m} P x_{ji}(t) \right) + N_0}. \tag{2.2}$$

Hence, the achievable data rate of CEU m using Shannon theorem is

$$R_m(t) = B \log_2 \left(1 + \beta \gamma_m(t)\right). \tag{2.3}$$

where is related to the target bit error rate (BER), and given by [8] as $\beta = -1.5/\ln(5\text{BER})$.

We consider two traffic types in our system, i.e. BE and VoIP. The sets of BE and VoIP users are denoted as $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively, with $\mathcal{M}_1 \cup \mathcal{M}_2 = \mathcal{M}$ and $\mathcal{M}_1 \cap \mathcal{M}_2 = \varnothing$. Based on the different service requirements for each type of user, two different utility functions are defined to represent their satisfaction.

For a BE user, the satisfaction is assumed to depend on its average throughput. Hence, the utility function of BE user m is defined as a monotonically increasing function of the average throughput $\bar{R}_m(t)$ at time t

$$U_m(t) = U_{BE,m}\left(\bar{R}_m(t)\right); \forall m \in \mathcal{M}_1, \tag{2.4}$$

where $\bar{R}_m(t)$ is estimated using an exponential filter as [9]

$$\bar{R}_m(t) = (1 - \rho_{BE})\bar{R}_m(t-1) + \rho_{BE} R_m(t). \tag{2.5}$$

The satisfaction of a VoIP user is assumed to depend on its service delay. Hence, the utility function for a VoIP user is defined as a monotonically decreasing function of its average queuing delay, which is given by

$$U_m(t) = U_{V,m}\left(\bar{d}_m(t)\right); \forall m \in \mathcal{M}_2, \tag{2.6}$$

where $\bar{d}_m(t)$ is its average queuing delay at time t. We estimate $\bar{d}_m(t)$ through the approach proposed in [10]. Define $Q_m(t)$ as the queue size in bits at the end of time slot t, and $\alpha_m(t)$ as the instantaneous arriving bits at the end of slot t. With departure rate $R_m(t)$, the queue size is calculated by

$$Q_m(t) = Q_m(t-1) - \min\{R_m(t)T_s, Q_m(t)\} + \alpha_m(t-1). \tag{2.7}$$

where T_s is the slot duration. Assuming that $Q_m(t)$ is ergodic, then with Little's Law the average delay can be estimated by $\bar{d}_m(t) = \bar{Q}_m(t)/\bar{\alpha}_m$, where $\bar{\alpha}_m$ denotes the time averaging arrival bits per slot, and $\bar{Q}_m(t)$ is the average queue size at time slot t. Similar to $\bar{R}_m(t)$ in (5), the average queue size is estimated by

$$\bar{Q}_m(t) = (1-\rho_V)\bar{Q}_m(t-1) - \rho_V Q_{V,m}(t). \tag{2.8}$$

This in turn leads to the estimate of the average delay via

$$\bar{d}_m(t) = (1-\rho_V)\bar{d}_m(t-1) + \rho_V \bar{\alpha}_m^{-1} Q_m(t). \tag{2.9}$$

Substitute (8) into (9), and the estimate of the average delay is ultimately expressed by

$$\begin{aligned}\bar{d}_m(t) =& (1-\rho_V)\bar{d}_m(t-1) + \rho_V \bar{\alpha}_m^{-1}[Q_m(t-1) \\ &- \min\{R_m(t)T_s, Q_m(t-1)\} + \alpha_m(t-1)].\end{aligned} \tag{2.10}$$

2.3 Utility-Based Joint Scheduling and Power Control

In this section, the maximum sum utility optimization problem is first formulated, and then a utility-based joint scheduling and power control algorithm is proposed for radio resource allocation.

2.3.1 Problem Formulation

Our objective is to maximize the system sum utility, so the objective function is formulated as

$$U(t) = \sum_{m_1 \in \mathcal{M}_1} U_{BE,m_1}\left(\bar{R}_{m_1}(t)\right) + \sum_{m_2 \in \mathcal{M}_2} U_{V,m_2}\left(\bar{d}_{m_2}(t)\right). \tag{2.11}$$

Note that $\bar{R}_{m_1}(t-1)$ and $\bar{d}_{m_2}(t-1)$ are known at time slot t. Hence, using Taylor expansion, to maximize (11) at time slot is equivalent to maximize

$$\Pi(t) = \sum_{m_1 \in \mathcal{M}_1} U'_{BE,m_1}\left(\bar{R}_{m_1}(t-1)\right)\bar{R}_{m_1}(t) + \sum_{m_2 \in \mathcal{M}_2} U'_{V,m_2}\left(\bar{d}_{m_2}(t-1)\right)\bar{d}_{m_2}(t). \tag{2.12}$$

Substitute (5) and (10) into (12), and let the CU control the service bit rate so that $R_{m_2}T_s \le Q_{m_2}(t-1)$. Then with $Q_{m_2}(t-1)$ and $\alpha_{m_2}(t-1)$ known at slot t, we have $\Pi(t)$ as a function only of $R_m(t)$.

$$\begin{aligned}\Pi(t) = &\sum_{m_1\in\mathcal{M}_1} \rho_{BE} U'_{BE,m_1}\left(\bar{R}_{m_1}(t-1)\right) R_{m_1}(t) \\ &- \sum_{m_2\in\mathcal{M}_2} \rho_V \bar{\alpha}_{m_2}^{-1} T_s U'_{V,m_2}\left(\bar{d}_{m_2}(t-1)\right) R_{m_2}(t).\end{aligned} \tag{2.13}$$

Note that the marginal utility functions $U'_{BE,m_1}(\cdot)$ and $U'_{V,m_2}(\cdot)$ are related to the scheduling weights or priorities, and thus play a key role in scheduling. Since $R_m(t)$ is related to $\mathbf{X}(t)$, $\Pi(t)$ turns out to be a function of $\mathbf{X}(t)$. Using $\Pi(\mathbf{X}(t))$ to represent $\Pi(t)$, (13) becomes

$$\Pi(\mathbf{X}(t)) = \sum_{m_1\in\mathcal{M}_1} \pi_{m_1} R_{m_1}(\mathbf{X}(t)) + \sum_{m_2\in\mathcal{M}_2} \pi_{m_2} R_{m_2}(\mathbf{X}(t)). \tag{2.14}$$

where π_{m_1} and π_{m_2} are fixed at time t, with

$$\begin{cases} \pi_{m_1} = U'_{BE,m_1}\left(\bar{R}_{m_1}(t-1)\right)\rho_{BE}, \\ \pi_{m_2} = -U'_{V,m_2}\left(\bar{d}_{m_2}(t-1)\right)\rho_V \bar{\alpha}_{m_2}^{-1} T_s. \end{cases} \tag{2.15}$$

Ultimately, the optimization problem is mathematically formulated as

$$\begin{aligned} &\max_{\mathbf{X}(t)} \Pi(\mathbf{X}(t)) \\ &\text{s.t. } 1) R_{m_2}T_s \le Q_{m_2}(t-1), \forall m_2 \in \mathcal{M}_2 \\ &\quad\;\; 2) \sum_{m=1}^{M} x_{nm}(t) \le 1, \forall n \in \mathcal{N}. \end{aligned} \tag{2.16}$$

That is, the optimization problem (16) becomes finding $\mathbf{X}^*(t)$ that maximizes $\Pi(\mathbf{X}(t))$ in (14) under the constraints 1) the instantaneous departure data size for VoIP users can be no more than the attainable waiting queue size, and 2) a BSS transmits to at most one CEU.

2.3.2 Algorithm Description

To solve the optimization problem in (16), a joint scheduling and power control algorithm is proposed in this section for the CoMP system.

We define the set of all feasible user schedules in each time slot as $\mathbb{X}(t)$. Assume $N = 3$, and CSI for all channels is perfectly known. Exploiting binary power control, then we have $(M+1)^3$ candidates in $\mathbb{X}(t)$ each time slot. In general, the complexity increases as the number of users per BSS increases. For a system with M users and N BSSs in the

Algorithm 1 Utility-Based Joint Scheduling and Power Control Algorithm

Initialization $P_n{=}P, \forall n{\in}\mathcal{N}$; $\bar{R}_{m_1}(0){=}0, \forall m_1{\in}\mathcal{M}_1$; $\bar{d}_{m_2}(0){=}0$, $\bar{Q}_{m_2}(0){=}0$, $\forall m_2{\in}\mathcal{M}_2$ at time slot $t{=}0$;
In each time slot t
  For each shedule $\mathbf{X}(t)$ in $\mathbb{X}(t)$
    Compute $\Pi\left(\mathbf{X}(t)\right)$ using (12)
  End
  $\mathbf{X}^*(t) = \arg\max_{\mathbf{X}(t)\in\mathbb{X}(t)} \Pi\left(\mathbf{X}(t)\right)$
  For each user $m_1 \in \mathcal{M}_1$
    Update $\bar{R}_{m_1}(t)$ using (5)
  End
  For each user $m_2 \in \mathcal{M}_2$
    Update $\bar{d}_{m_2}(t)$, $\bar{Q}_{m_2}(t)$ using (10), (8)
  End
  Advance t
End

cluster, the complexity of the proposed joint scheduling and power control algorithm is $O\left((M+1)^N\right)$.

The algorithm starts with an empty user set, and assigns each BSS with the same maximum transmit power P. Then in each time slot, the algorithm does the exhaustive search of all the feasible user schedules in the set $\mathbb{X}(t)$ for the optimal user group $\mathbf{X}^*(t)$ that gives the largest $\Pi\left(\mathbf{X}(t)\right)$. At the end of each time slot t, $\bar{R}_m(t)$, $\bar{d}_m(t)$, and $\bar{Q}_m(t)$ are updated based on $\mathbf{X}^*(t)$. The algorithm is outlined in Algorithm 1.

2.4 Simulation Results

We focus on the downlink of flat fading channel in a cellular system, with carrier frequency of 2GHz. Assume path loss with $L(d) = 128.1{+}37.6\log_{10} d$ in dB [4], where is the distance in km, and log-normal shadowing with zero-mean and 10 dB standard deviation. A cluster of three BSSs is taken into account. The cell radius is set to 500m. A number of users are uniformly allocated in the cell-edge area of the CoMP cluster, where the long term channel gain is under the threshold -100 dB. 1000 independent trials are evaluated by Monte-Carlo simulation under various numbers of cell-edge users per cluster.

We randomly drop the two different types of users in the CoMP network, i.e., BE users and VoIP users, with the probability of 50% for each. The target BER for data transmission is prescribed as 10^{-5}. We then assume the full-buffer traffic model for BE users with the utility function defined as

$$U_{BE,m_1}\left(\bar{R}_{m_1}(t)\right) = \ln\left(\bar{R}_{m_1}(t)\right). \tag{2.17}$$

For the VoIP users we consider a VoIP traffic model with packet inter-arrival time of 5ms, and packet size of 10 bytes. Regarding QoS requirements, we set the maximum queuing delay to be 15ms. VoIP calls for users experiencing packet delays greater than the maximum delay result in call outage, and these users are redropped at another allocation. Hence, the utility function is defined as

$$U_{V,m_2}\left(\bar{d}_{m_2}(t)\right) = -\frac{\log_{10}\delta_{m_2}}{2\check{d}_{m_2}}\left(\check{d}^2_{m_2} - \bar{d}^2_{m_2}(t)\right), 0 < \delta_{m_2} < 1, \tag{2.18}$$

where $\check{d}_{m_2}$ is the maximum allowable queuing delay for VoIP users; $\delta_{m_2} = 0.1$ is a constant, chosen to balance the priorities of different types of users [11].

Recall (13), the best-effort users with lower average throughput can get higher priority in the scheduling if $U_{BE,m_1}(\cdot)$ is chosen as in (17). Similarly, with $U_{V,m_2}(\cdot)$ defined as in (18), the VoIP users will gain higher priorities if they experience larger delays. In fact, the marginal utility function of (18) turns out to be the largest-weighted-average-delay-first (LWADF) scheduling [5], i.e., users in the queue experiencing the largest average delay have the highest priorities and should be served first in each round of scheduling. Besides, ρ_{BE} and ρ_V also play an important role in balancing priorities of the two types of users in scheduling, and are prescribed as $\rho_{BE} = 0.01$ and $\rho_V = 0.05$ respectively.

The average sum utility is evaluated as the assessment of the proposed utility-based joint scheduling and power control algorithm, named C-UBPC. Meanwhile, as a special case of the proposed utility-based joint scheduling and power control algorithm, another algorithm with the power of all BSSs in the cluster always turned on, named C-UB, is also considered and assessed. Jain's Fairness Index (JFI) [4] of utilities is investigated as a fairness measure of user satisfaction based on users' average utility

$$\mathrm{FI} = \frac{\left(\sum_{m=1}^{M}\bar{U}_m\right)^2}{M\left(\sum_{m=1}^{M}\bar{U}^2_m\right)}. \tag{2.19}$$

Furthermore, we also show the resulting average call outage ratio for the VoIP traffic, and the average user throughput for the BE users, respectively. As comparison, three other algorithms are considered:

1. *Coordinated proportional-fair scheduling without power control (C-PF)*: The algorithm is aimed to maximize the proportional throughput-fair index [5] with joint transmission, but the differentiations of diverse traffic models are not considered.

2. *Utility-based scheduling without joint transmission or power control (NC-UB)*: Similar to the proposed C-UB algorithm, but no joint transmission is supported.

3. *Proportional-fair scheduling without joint transmission or power control (NC-PF)*: Similar to 1 but no joint transmission is supported either.

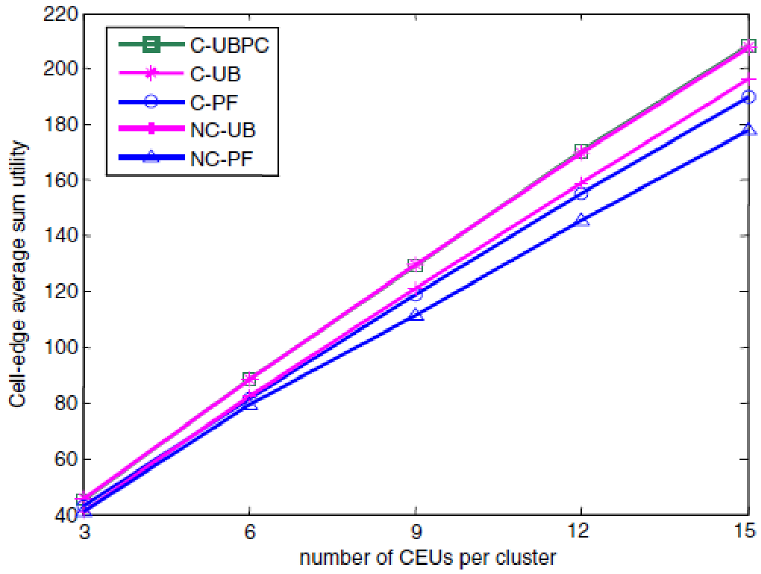


Figure 2.2: Average sum utility vs. number of CEUs per cluster.

Fig.2 shows the cell-edge average sum-utility of the algorithms considered in this paper with respect to the number of cell-edge users per cluster. It can be seen that with joint transmission the proposed C-UBPC and C-UB algorithms achieve the highest aggregate utility. Additionally, even without joint transmission, the NC-UB algorithm still achieves slightly better performance than C-PF algorithm by exploiting different utility functions. The C-UBPC algorithm with binary power control achieves similar performance as the C-UB algorithm. However, there is a power saving in the C-UBPC algorithm based on the simulation results, with the average number of BSS turned on as 1.96-2.46; while in the C-UB algorithm, the number of BSS turned on is always 3. The NC-PF algorithm yields the lowest utility given neither joint transmission nor utility differentiations.

To further improve our understanding of joint transmission and utility-based scheduling in this paper, in Fig.3, Fig.4, we plot the average throughput for BE users, and the average VoIP call outage ratio, respectively. We can see that as the traffic gets heavier, for all the algorithms the average BE throughput decreases, and VoIP call outage increases within the prescribed QoS call outage, respectively. Nonetheless, under all the traffic conditions, by taking advantage of good diverse QoS provisioning through exploiting utility functions, the utility-based algorithms achieve relatively better performance than the proportional fair algorithms, i.e., the C-PF and NC-PF algorithms.

As seen from Fig.3 and Fig.4, compared with the NC-UB algorithm with no joint transmission supported, the proposed C-UBPC and C-UB algorithms achieve higher average BE throughput, and meanwhile, improve the VoIP service with much lower average call outage. From Fig.3, we can see that with binary power control, the C-UBPC algorithm has slightly better performance in terms of the average BE throughput while achieving a

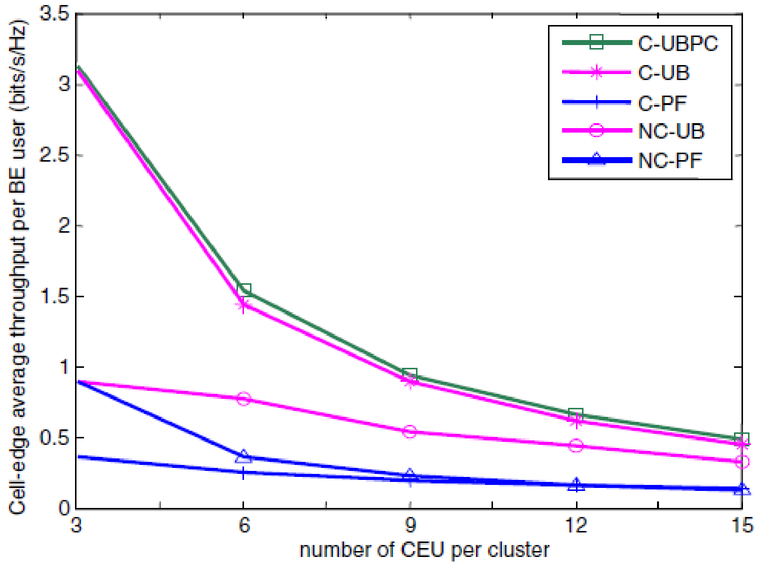


Figure 2.3: Average throughput per BE user vs. number of CEUs per cluster.

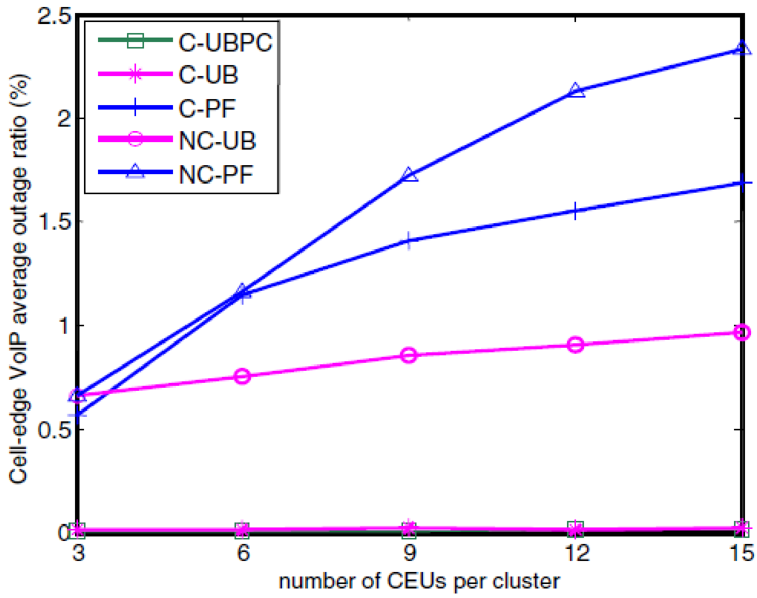


Figure 2.4: Average VoIP outage ratio vs. number of CEUs per cluster.

better power saving, compared to the C-UB algorithm without power control.

The utility JFIs of the five algorithms are plotted in Fig.5. It shows that with diverse utility functions and joint transmission, the proposed C-UBPC and C-UB algorithms achieve the best user utility fairness. By utilizing joint transmission, the C-PF algorithm also achieves higher utility fairness by contrast to the NC-UB and NC-PF algorithms.

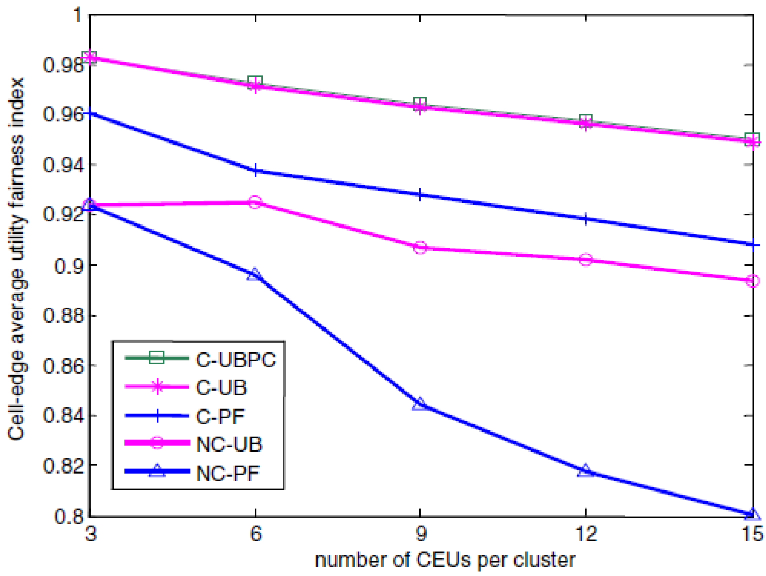


Figure 2.5: Utility JFI vs. number of CEUs per cluster.

2.5 Conclusions

In this paper, we consider the downlink of a CoMP cluster with three neighboring BSSs. A utility-based joint scheduling and power control algorithm is proposed in order to maximize the cell-edge sum utility of the CoMP system with mixed VoIP and best-effort traffic. First, we mathematically formulate the objective function with respect to the average throughput and queuing delay for joint transmission and power control. Then a resource allocation algorithm is developed to jointly assign a group of users in the cluster. The simulation results demonstrate that the proposed algorithm provides a significant improvement in terms of system sum utility and user fairness. The results also show that the algorithm increases the average throughput of best effort user by greater than 45% and decreases the average call outage of VoIP user by greater than 98%, compared to traditional scheduling schemes without joint transmission and diverse QoS provisioning.

The results in this paper focus on flat fading channel in a single cluster, but we do not consider the complexity introduced by joint scheduling. In future work, joint scheduling and power control problems in multi-subchannel systems with multi-cluster interference will be addressed, as well as less complex algorithms.

References

[1] 3GPP TR 36.814, V9.0.0, Further advancements for EUTRA physical Layer Aspects, Mar. 2010.

[2] L. Venturino, N. Prasad and X. Wang, Coordinated scheduling and power allocation in downlink multicell OFDMA networks, IEEE Trans. on Veh. Technol., vol. 58, no. 6, pp. 2835-2848, Jul. 2009.

[3] L. Chen, W. Chen, X. Zhang, Y. Zhang, C. Xiong and D. Yang, Intercell coordinated resource allocation for mobile WiMAX system, Proc. IEEE WCNC, pp. 1-6, Apr. 2009.

[4] J. Li, X. Xu, X. Chen, X. Tao, H. Zhang, T. Svensson, and C. Botella, Downlink radio resource allocation for coordinated cellular OFDMA networks, IEICE Trans. Commun, vol.e93-b, no. 12, pp. 3480-3488, 2010.

[5] X. Wang, G. B. Giannakis, and A. G. Marques, A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks, Proceedings of the IEEE, vol. 95, no. 12, pp. 2410-2431, Dec. 2007.

[6] J. Puttonen, N. Kolehmainen, T. Henttonen, M. Moisio, and M. Rinne, Mixed traffic packet scheduling in UTRAN Long Term Evolution Downlink, in Proc. PIMRC'08, pp.1-5.

[7] M. Katoozian, K. Navaie, and H. Yanikomeroglu, Utility-based adaptive radio resource allocation in OFDM wireless networks with traffic prioritization, IEEE Trans. Wireless Commun., vol. 8, no. 1, 2009.

[8] A. J. Goldsmith and S.-G. Chua, Variable-rate variable-power MQAM for fading channels, IEEE Trans. Commun., vol. 45, no. 10, pp. 1218- 1230, Oct. 1997.

[9] G. Song and Y. Li, Cross-layer optimization for OFDM wireless networks-part II: algorithm development, IEEE Trans. on Wireless Commun., vol.4, no.2, pp. 625-634, Mar. 2005.

[10] G. Song and Y. Li, Cross-layer resource allocation and scheduling in wireless multi-carrier networks, Ph.D. dissertation, Georgia Inst. Technol., Apr. 2005.

[11] M. Andrews, K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, and P. Whiting, Providing quality of service over a shared wireless link, IEEE Commun. Mag., vol. 39, no. 2, pp. 150-154, Feb. 2001.

PAPER B

Joint Scheduling for Multi-Service in Coordinated Multi-Point OFDMA Networks

Binru Huang, Jingya Li, Tommy Svensson

To be published in
IEEE 75th Vehicular Technology Conference (VTC Spring)

Yokohama, Japan, 2012

Abstract

In this paper, the issues upon user scheduling in the downlink packet transmission for multiple services are addressed for coordinated multi-point (CoMP) OFDMA networks. We consider mixed traffic with voice over IP (VOIP) and best effort (BE) services. In order to improve cell-edge performance and guarantee diverse quality of service (QoS), a utility-based joint scheduling algorithm is proposed, which consists of two steps: ant colony optimization (ACO) based joint user selection and greedy subchannel assignment. We compare the proposed algorithm with the opimal algorithm and the greedy user selection (GUS) based scheme. Via simulation results, we show that the performance of our proposed algorithm well approaches to that of the optimal solution. It is also observed that 95% of BE users are satsified with average cell-edge data rate greater than 200kbps by using either of the two algorithms. Whereas, our proposed algorithm ensures that more than 95% of VoIP users are satisfied with packet drop ratio less than 2%, compared to 78% by the GUS based algorithm.

Keywords: Ant colony optimization (ACO), coordinated multi-point (CoMP), OFDMA, multiple services, scheduling

3.1 Introduction

Future wireless networks are required to satisfy not only high-speed data applications, but also delay-sensitive applications such as voice over IP (VoIP). To support broadband and multimedia services, orthogonal frequency division multiple access (OFDMA) is identified as a leading candidate for future wireless networks [1]. Additionally, full frequency reuse is envisioned to offer high spectrum efficiency in wireless networks to guarantee diverse service requirements. However, intercell interference (ICI) due to the universal reuse of spectral resources can significantly degrade the system performance and quality of service (QoS).

Coordinated Multi-Point (CoMP) joint transmission is proposed as a promising technique to mitigate ICI and improve system spectral efficiency [2]. In CoMP joint transmission systems, coordinated base stations (CBSs) are connected via a high-speed backbone, and ICI is significantly mitigated by applying the signals transmitted from other CBSs to assist the transmission. In order to make the inter-BS communication overhead affordable, user grouping, e.g., serving subsets of terminals with CoMP joint transmission [3], and clustering of BSs, i.e., dividing the network into small clusters of BSs [4], have been considered. Currently, studies on radio resource management (RRM) cooperation for CoMP systems mainly focus on single service networks, assuming full buffer for all the users [4]-[6]. A convex optimization based coordinated subchannel assignment scheme is proposed in [4] for coordinated OFDMA networks with the objective of maximizing total cell-edge utility. The combinatorial optimization problem is then decomposed into independent optimization problems for each subchannel. The approach proposed in [5] reduces the complexity of centralized multi-cell scheduling by exploiting the idea of greedy user selection (GUS), however, at the expense of degradation in the system performance. In [6], a two-phase scheduling method is proposed considering joint transmission, where user grouping is solved via graph theory in the first phase and subchannel allocation is accomplished in the second phase.

Notice that, [4]-[6] assume only single service for all the users in the networks. In [7] multiple services are taken into account for designing the scheduling algorithms with CoMP joint transmission, under the assumption of a flat fading channel. In this paper, the joint scheduling problem is addressed for multi-cell multi-subchannel scenarios. The system is divided into several disjoint clusters, where each cluster performs joint scheduling independently. Users in each cluster are divided into two groups, i.e., VoIP users and best effort (BE) users according to their traffic patterns. Different utility functions are defined for VoIP users and BE users with respect to different individual QoS requirements. The objective is to maximize the total utility in each cluster.

In general, to solve the scheduling problem in multi-cell multi-service OFDMA systems with CoMP joint transmission is computationally intensive, especially when the number of users and subcarriers are large. Ant colony optimization (ACO) has recently attracted growing attention in the studies of complex combinatorial optimization problems in com-

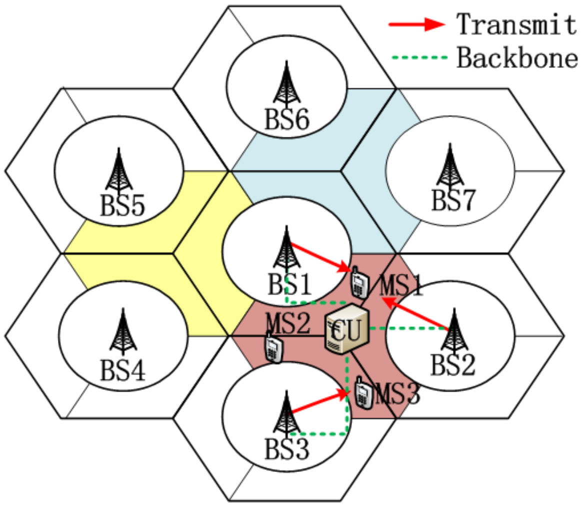


Figure 3.1: System model of clustered CoMP networks

munications networks [8]-[9]. Based on ACO, in this paper a joint user selection scheme is proposed for each subchannel as a first step. Then, subchannel assignment is based on the idea of greedy algorithm. Via simulation results, the proposed algorithm is shown to yield sum utility which well approaches that of the optimal solution, and could achieve better cell-edge performances compared to the GUS based algorithm.

The rest of the paper is organized as follows. In Section II we briefly describe the system model and formulate the optimization problem. In Section III we introduce the proposed joint scheduling algorithm, involving two steps: ACO joint user selection and greedy subchannel assignment. The simulation results are presented in Section IV, and conclustions are given in Section V.

3.2 System Model and Problem Formulation

3.2.1 Clustered CoMP Networks with Multi-Services

We consider the downlink of a cellular OFDMA system, in which CoMP joint transmission is supported. Assume a static clustering technique, where all the base station sectors (BSSs) in the network are divided into a number of disjoint clusters of coordinated BSSs. As in Fig.3.1), each cluster consists of three neighboring BSSs, and is marked by different color respectively. BSSs within the same cluster are connected to a central unit (CU) through high-speed backbones. According to the long term channel gain, mobile stations (MSs) are divided into two classes, namely cell-center MSs and cell-edge MSs. Only MSs at the cell-edge area are considered in this paper. For simplicity, we assume single antenna for both transmitter and receiver. A spectrum bandwidth B is shared among all the BSSs with universal frequency reuse of 1, and is diveded into K subchannels, each with a bandwidth of $B_{SCH} = B/K$. The maximum transmit power for each BSS is P, which is equally distributed over the subchannels.

Assume each MS has an individual queue to receive its incoming packets, and each CU performs joint scheduling independently according to instantaneous feedback queue information (QI) and channel state information (CSI). The OFDM signaling is time-slotted with the duration of each time slot as T_s. Since joint scheduling is performed independently in each cluster, thus to simplify notation, we focus on a single cluster consisting of a BSS set $\mathcal{N}$ and a cell-edge MS set $\mathcal{M}$, with the cardinality $|\mathcal{N}| = N$, $|\mathcal{M}| = M$ respectively. For any given time slot t, let P_n^k denote the transmit power of BSS n on subchannel k, and $G_{m,n}^k$ denote the channel gain between MS m and BSS n on subchannel k. Suppose each subchannel of a BSS can only be assigned to one MS. $x_{m,n}^k$ indicates whether subchannel k of BSS n is assigned to MS m, and thus we have

$$x_{n,m}^k = \begin{cases} 1, \text{ BSS}\, n \text{ transmit to MS}\, m \text{ on subchannel}\, k; \\ 0, \text{ otherwise;} \end{cases} \tag{3.1}$$

subject to $\sum_{m=1}^{M} x_{n,m}^k = 1$. The joint transmission BSS set for MS m on subchannel k can be denoted by $S_m^k = \{n | x_{n,m}^k = 1, n \in \mathcal{N}\}$. Hence, given the power N_0 of the additive white Gaussian noise (AWGN), the signal-to-interference-and-noise ratio (SINR) for MS m on k^{th} subchannel with non-coherent joint transmission is given by

$$\gamma_m^k = \frac{\sum_{n \in \mathcal{N}} P_n^k G_{m,n}^k x_{m,n}^k}{I_m^k + N_0}, \tag{3.2}$$

where I_m^k is the cochannel ICI on subchannel k for MS m including intra-cluster interference and inter-cluster interference,

$$I_m^k = \sum_{i \in \mathcal{N}} P_i^k G_{m,i}^k \left(\sum_{\substack{s \in \mathcal{M} \\ s \neq m}} x_{s,i}^k \right) + \sum_{j \notin \mathcal{N}} P_j^k G_{m,j}^k. \tag{3.3}$$

Thus the achievable bit rate of MS m on subchannel k is expressed as

$$R_m^k = B_{SCH} \log_2(1 + \beta \gamma_m^k), \tag{3.4}$$

where β is a constant related to the target bit error rate (BER) given by $\beta = -1.5/\ln(5\,\text{BER})$ [4]. The instantaneous data transmission rate for MS m at time slot t becomes

$$R_m(t) = \sum_{k=1}^{K} R_m^k. \tag{3.5}$$

Based on different traffic types, the M users in the cluster are classified into two subsets, i.e., BE and VoIP users. We assume full buffer for BE users, with a prescribed minimum average bit rate $\bar{R}_{\min}$. For delay sensitive VoIP services, we consider very bursty and low bit rate, and impose a maximum allowed instantaneous queueing delay $\tau_{\max}$, i.e., $\tau_m(t) \leq \tau_{\max}$, where $\tau_m(t)$ denotes the instantaneous packet delay of user m. The expired packets with latency larger than $\tau_{\max}$ will be dropped at the transmitter. Besides, at the receiver any unsucessfully received packet that fails to be decoded is also discarded. The packet drop ratio should be maintained at an acceptable level to guarantee QoS for VoIP services.

3.2.2 Optimization Objective

Let $\mathcal{M}_B$ and $\mathcal{M}_V$ denote the subsets of BE and VoIP users, respectively. According to different QoS requirements of different services, we utilize two different utility functions to represent satisfaction levels for the two different types of users. For a BE user, user's satisfaction increases as the average data rate increases, and thus the utility is defined as a concave and monotonically increasing function of its average data rate, i.e., $U_m(t) = U_{B,m}(\bar{R}_m(t)); \forall m \in \mathcal{M}_B$, where $\bar{R}_m(t)$ is the average data rate of BE user m at time slot t, and is estimated using an exponential filter [4]

$$\bar{R}_m(t) = (1 - \rho_B)\bar{R}_m(t-1) + \rho_B R_m(t). \tag{3.6}$$

We consider the utility of VoIP users as a monotonically decreasing function of the average delay, i.e., $U_m(t) = U_{V,m}(\bar{\tau}_m(t)); \forall m \in \mathcal{M}_V$, where $\bar{\tau}_m(t)$ is the average delay of VoIP user m at time slot t, and expressed as in [7] by

$$\begin{aligned} \bar{\tau}_m(t) =& (1 - \rho_V)\bar{\tau}_m(t-1) + \rho_V \bar{\alpha}_m^{-1}[Q_m(t-1) \\ & - \min\{R_m(t)T_s, Q_m(t-1)\} + \alpha_m(t-1)], \end{aligned} \tag{3.7}$$

given that $\bar{\alpha}_m$ denotes the time average arriving bits per slot, $Q_m(t–1)$ is the instantaneous queue size at time slot $t-1$, and $\alpha_m(t-1)$ denotes the number of instantaneous arriving bits at the end of time slot $t-1$.

Our objective is to maximize sum utility of the MSs in each disjoint cluster. Therefore, the objective function is

$$U(t) = \textstyle\sum_{m_1 \in \mathcal{M}_B} U_{B,m_1}\big(\bar{R}_{m_1}(t)\big) + \sum_{m_2 \in \mathcal{M}_V} U_{V,m_2}(\bar{\tau}_{m_2}(t)). \tag{3.8}$$

Note that $\bar{R}_{m_1}(t-1)$ and $\bar{\tau}_{m_1}(t-1)$ are fixed at time slot t. Using Taylor expansion, (3.13) can be approximated as [4]

$$\begin{aligned} U(t) = & \textstyle\sum_{m_1 \in \mathcal{M}_B} U'_{B,m_1}\big(\bar{R}_{m_1}(t-1)\big)\bar{R}_{m_1}(t) \\ & + \textstyle\sum_{m_2 \in \mathcal{M}_V} U'_{V,m_2}(\bar{\tau}_{m_2}(t-1))\bar{\tau}_{m_2}(t). \end{aligned} \tag{3.9}$$

Control the departure rate so that $R_{m_2}(t)T_s \leq Q_{m_2}(t-1)$. Substituted with (3.6) and (3.7), thus to maximize (3.9) is equivalent to maximize

$$\Pi(t) = \textstyle\sum_{m \in \mathcal{M}} \omega_m R_m(t), \tag{3.10}$$

where

$$\omega_m = \begin{cases} \rho_B U'_{B,m}\big(\bar{R}_m(t-1)\big), & m \in \mathcal{M}_B; \\ -\rho_V \bar{\alpha}_m^{-1} U'_{V,m}(\bar{\tau}_m(t-1)), & m \in \mathcal{M}_V. \end{cases} \tag{3.11}$$

Accordingly, the optimization objective turns out to be a linear function in terms of $R_m(t)$.

3.2.3 Problem Formulation

Let $\mathbf{X} = \left[x_{m,n}^k\right]$ be the subchannel assignment matrix. Since $R_m(t)$ is decided by $\mathbf{X}$, substituting (3.5) we thus formulate the optimization problem in the coordinated OFDMA system as

$$\begin{aligned} &\max_{\mathbf{X}} \textstyle\sum_{m\in\mathcal{M}} \sum_{k=1}^{K} \omega_m R_m^k, \\ &\text{s.t. } 1) \textstyle\sum_{k=1}^{K} R_m^k T_s \le Q_m, \forall m \in \mathcal{M}_V, \\ &\quad\; 2) \textstyle\sum_{m=1}^{M} x_{m,n}^k = 1, \\ &\quad\; 3) \tau_m(t) \le \tau_{\max}, \forall m \in \mathcal{M}_V. \end{aligned} \tag{3.12}$$

That is, the optimization problem becomes to decide $\mathbf{X}$ such that (3.11) is maximized, subject to the following constraints 1) the CU does not waste spectrum by serving an empty queue or a queue with less bits than achivable departure rate, 2) each subchannel of a BSS can only be assigned to one MS, and 3) the instantaneous latency of a VoIP packet should be no greater than the maximum allowed delay.

3.3 Proposed Joint Scheduling Algorithm

The optimization problem presented in (3.12) is an NP-hard combinatorial problem in this paper. ACO is introduced to solve the problem. ACO was inspired by the way that real ants find the shortest path in search for food. Ants deposit some amount of pheromone on the path to the food. More ants passing a way to the food place, the higher amount of pheromone the path will have, and more attractive that path will become. Based on this idea, ACO is proposed to solve difficult tasks and figure out combinatorial problems [8].

In an ant colony system, a number of edges and vertices make all the possible paths from the nest to the food source. The ants make decision in parallel and asynchronously choose the next edge in all vertices. The optimal solution would be a list of edges and vertices which makes the shortest route between the nest and food. Virtual ants choose their paths using two factors: *visibility* and *trail intensity*. Visibility is what an ant can see, and usually it is considered as the inverse of the edge length. Trail intensity is the pheromone left at each vertex, and also related to the number of ants passing through the edge [9].

Based on the idea of ACO, we propose a two-step joint scheduling algorithm in this section. We firstly model the joint user selection on each subchannel as an ACO problem. Then all the subchannels are allocated for each group of coordinated BSSs by using greedy algorithm.

3.3.1 ACO Based Joint User Selection

We model the combinatorial user selection problem on each subchannel into a pseudograph with each vertex representing a BSS in the cluster. A set of ants $\mathcal{A}$ is considered in our ant colony system. A combination of BSS and user utility is denoted by an edge, which is connected to a vertex. Edges are associated with the user utility augment or priority, which indicates how likely an edge will be selected. Hence, it is reasonable to design the visibility as

$$V_{n,m}^{k} = \begin{cases} \omega_m R_{n,m}^{k}, & Q_m > 0; \\ 0, & \text{otherwise}; \end{cases} \tag{3.13}$$

where $R_{n,m}^{k}$ denotes the data rate transmitted from BSS n to MS m over subchannel k, and is estimated using (3.4) based on the local SINR approximation considering only inter-cluster interference

$$\gamma_{n,m}^{k} = \frac{P_n^k G_{m,n}^{k} x_{m,n}^{k}}{\sum_{l \notin \mathcal{N}} P_l^k G_{m,l}^{k} + N_0}. \tag{3.14}$$

Then the probability of selecting an edge to a vertex is given by [9]

$$p_{n,m}^{k} = \frac{[T_{n,m}^{k}]^{\alpha} \cdot [V_{n,m}^{k}]^{\beta}}{\sum_{m=1}^{M} [T_{n,m}^{k}]^{\alpha} \cdot [V_{n,m}^{k}]^{\beta}}, \tag{3.15}$$

where $T_{n,m}^{k}$ is the trail intensity of MS m receiving service from BSS n over subchannel k, and is initially set to 1. Trail intensity shows the popularity of an edge. Hence, the trail intensity is updated in each iteration by

$$T_{n,m}^{k}(l+1) = \xi T_{n,m}^{k}(l) + \Delta T_{n,m}^{k}, \tag{3.16}$$

where $0 < \xi < 1$ is the pheromone evaporation parameter, and $\Delta T_{n,m}^{k}$ is calculated by

$$\begin{aligned} \Delta T_{n,m}^{k,a} &= \begin{cases} \kappa / L_a, & a^{\text{th}} \text{ ant select } m \text{ to vertex } n; \\ 0, & \text{otherwise}; \end{cases} \\ \Delta T_{n,m}^{k} &= \textstyle\sum_{a \in \mathcal{A}} \Delta T_{n,m}^{k,a}, \end{aligned} \tag{3.17}$$

given that κ is a constant, and L_a^k is the tour length of a^{th} ant which corresponds to the total utility augment subject to the rate control constraint 1) in (3.12)

$$\Delta L_{n,m}^{k,a} = \begin{cases} \omega_m R_{n,m}^k, & R_{n,m}^k T_s \leq Q_m, m \in \mathcal{M}_V; \\ 0, & \text{otherwise}; \end{cases} \qquad (3.18)$$
$$L_a^k = \textstyle\sum_{n=1}^N \Delta L_{n,m}^{k,a},$$

By substituting (3.13) into (3.15), we can see that a user with empty data queue will never get served, which makes sure that the spectrum is not wasted serving a empty queue.

3.3.2 Greedy Subchannel Assignment

After performing ACO joint user selection on each subchannel, the queue size of each VoIP user requires adjustment such that the spectrum is not wasted serving a VoIP user redundantly. Hence, the queue size estimate of VoIP user m after multi-cell scheduling on subchannel k is updated successively as

$$\hat{Q}_m^{(k)} = \hat{Q}_m^{(k-1)} - R_m^k T_s \textstyle\sum_{n=1}^N x_{m,n}^k,\ m \in \mathcal{M}_V. \qquad (3.19)$$

In turn, the visibilities and user priorities should be reassigned according to $\hat{Q}_m^{(k)}$ on the remaining available subchannels. The proposed joint scheduling algorithm is presented in Algorithm 1.

3.3.3 Computational Complexity

A number of ants as well as iterations are in need for ACO joint user selection to improve the quality of solutions and acquire a final solution for joint scheduling. Therefore, the computational complexity of the proposed ACO joint user selection strategy could be much greater than the GUS scheme proposed in [5]. However, the ACO joint user selection can be implemented in parallel on multi-processors, which makes it an efficient algorithm for practical use. As long as there is a constant positive lower bound for the pheromone, ACO algorithm has been proved to be able to converge to the optimal solution [10], though the time for convergence is not given. Hence, considering simulation feasibility, the number of iterations can be predefined. It is shown that ACO based scheduling strategy can still achieve better performances than other suboptimal algorithms even without convergence [11].

3.4 Simulation Results

The simulation is conducted in a multi-cell OFDMA network with cell radius of 500 m. The total bandwidth is 3 MHz with the carrier frequency of 2 GHz and 15 parallel subchannels. The network is divided into multiple static disjoint clusters, with each cluster consisting of three adjacent BSSs. The propagation model takes into account path loss and fast fading, with the path loss as $L(d) = 128.1 + 37.6 \log_{10} d$ in dB [4], where d is the distance in km. A number of users are uniformly allocated with 50% of VoIP users and 50% of BE users in the cell-edge area, where the long term channel gain is assumed under the threshold -100 dB. The target BER for data transmission is prescribed as 10^{-5}. In the step of ACO joint user selection, 20 iterations are run in total with 10 ants used per iteration [8]. α and β are both set to 1, and the evaporation coefficient ξ and constant κ are set to 0.9 and 1 respectively as in [11]. Inter-cluster interference is taken into account. 1000 independent trials are evaluated by Monte-Carlo simulation under various numbers of cell-edge users per cluster.

Barrier Function is used to model the utility function for BE users [12]

$$U_{B,m_1}(\bar{R}_{m_1}) = \ln(\bar{R}_{m_1}) + (1 - e^{-\varphi(\bar{R}_{m_1} - \bar{R}_{\min})}), \tag{3.20}$$

where $\bar{R}_{\min}$ denotes the minimum target bit rate, assumed to be 200 Kbps. φ determines how aggressively the utility increases when $\bar{R}_{m_1}$ approaches $\bar{R}_{\min}$, and is set to 0.02. For the VoIP users, we model the voice traffic by an ON/OFF data arrival process with packet inter-arrival time of 10 ms. The full-rate voice frame size is 160 bits in ON period, and no transmit data otherwise. The maximum queuing delay $\tau_{\max}$ is 25 ms, and the required maximum packet drop ratio δ_{m_2} is 2%. A packet experiencing delay larger than $\tau_{\max}$ will be discarded. Hence, the utility function for the VoIP users is defined as

$$U_{V,m_2}(\bar{\tau}_{m_2}) = -\frac{\log_{10} \delta_{m_2}}{2\tau_{\max}} (\tau_{\max}^2 - \bar{\tau}_{m_2}^2), 0 < \delta_{m_2} < 1. \tag{3.21}$$

In fact, the marginal function of (3.21) turns out to be the modified-largest-weighted-delay-first (M-LWDF) utility [7].

To evaluate the performance of the proposed joint scheduling algorithm (shortly written as ACO), a two-step GUS based algorithm (referred as GUS) is given for comparison. In GUS algorithm, we firstly apply greedy user selection on each subchannel. Then the subchannel assignment is performed in the same way as the second step of our proposed algorithm. The performance of optimal user selection by exhaustive search (referred as Optimal) is also evaluated for purpose of comparison. Considering simulation feasibility, only the cases with small number of users are simulated for the Optimal algorithm.

The cell-edge sum utilities of the three algorithms are plotted in Fig.3.2a). It shows that both ACO and GUS algorithm yield a sum utility very close to Optimal algorithm, however, with much lower complexity. For a better observation, the sum utility increment ratio of ACO over GUS instead of sum utilities, is plotted in Fig.3.2b) in the cases when the number of cell-edge users is greater than 15. It shows that ACO yields higher cell-edge sum utility than GUS. The utility increment is small but it increases as the number of cell-edge users increases. Accordingly, we also plot in Fig.3.3 the VoIP packet drop ratio

Algorithm 2 Proposed Joint Scheduling Algorithm

Initialize $P_n^k = P/K, \forall n \in \mathcal{N}$; $\hat{Q}_m^{(1)} = Q_m(t-1), m \in \mathcal{M}_V$; $R_m^k = 0, \forall m \in \mathcal{M}, \forall k \in \mathcal{K} = \{1,2,...,K\}$
for $k = 1$ to K **do**
Step1: ACO based coordinated scheduling
  Initialize $T_{n,m}^k = 1$, $\Delta T_{n,m}^k = 0$, $\forall m \in \mathcal{M}, \forall n \in \mathcal{N}$
  Calculate $V_{n,m}^k$, and $p_{n,m}^k$, $\forall m \in \mathcal{M}, \forall n \in \mathcal{N}$
  for $l = 1$ to the number of iterations **do**
    for all $a \in \mathcal{A}$ **do**
      for $n = 1$ to N **do**
        Select an edge m to visit a new vertex n with probability of $p_{n,m}^k$
        Update $\Delta T_{n,m}^k$ according to selected m and n
      end for
    end for
    Save the best solution so far
    Update $T_{n,m}^k$, and $p_{n,m}^k$, $\forall m \in \mathcal{M}, \forall n \in \mathcal{N}$
  end for
Step2: Greedy subchannel assignment
  Assign the subchannel k according to the best solution
  Update $\hat{Q}_m^{(k)}$, $\forall m \in \mathcal{M}$
end for

and average user throughput of BE users in the cases when the number of users is larger than 15.

Fig.3.3a) shows the VoIP packet drop ratio of the two algorithms. We can see that the two algorithms both achieve low packet drop ratio and satisfy the prescribed QoS requirement, but by contrast, ACO outperforms GUS by achieving much lower packet drop ratio. The average user throughputs of cell-edge BE users are plotted in Fig.3.3b). It shows that average user throughput decreases as the number of cell-edge users increases. QoS is satisfied in all the cases for both ACO and GUS, with GUS achieving higher average user throughput than ACO. However, when traffic gets heavier, the cell-edge average user throughput of ACO approaches to that of GUS.

The user satisfaction ratio, i.e., the ratios of the numbers of QoS-satisfied users to the total number of cell-edge users of the two algorithms, are also plotted for both BE and VoIP users in Fig.3.4. In the plotting, we illustrate that more than 95% of the BE users are satisfied utilizing either ACO or GUS when the number of cell-edge users is within 27. For VoIP users, more than 95% of the users are satisfied for all the cases by exploiting ACO. However, using GUS only about 78% of the cell-edge users are satisfied. Hence, we can conclude that ACO yields higher sum utility at the cell-edge area, and in turn achieves better performance in balancing average throughput of BE users and packet drop ratio of VoIP users.

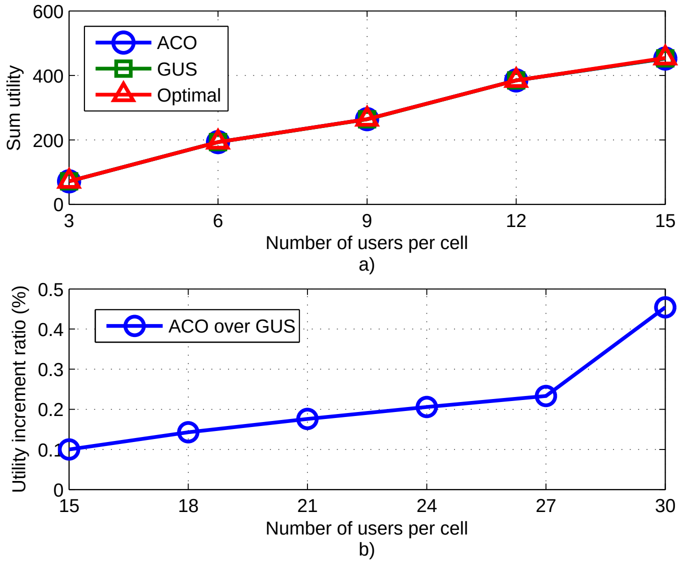


Figure 3.2: a) Cell-edge sum utility b) Cell-edge sum utility increment ratio of ACO over GUS

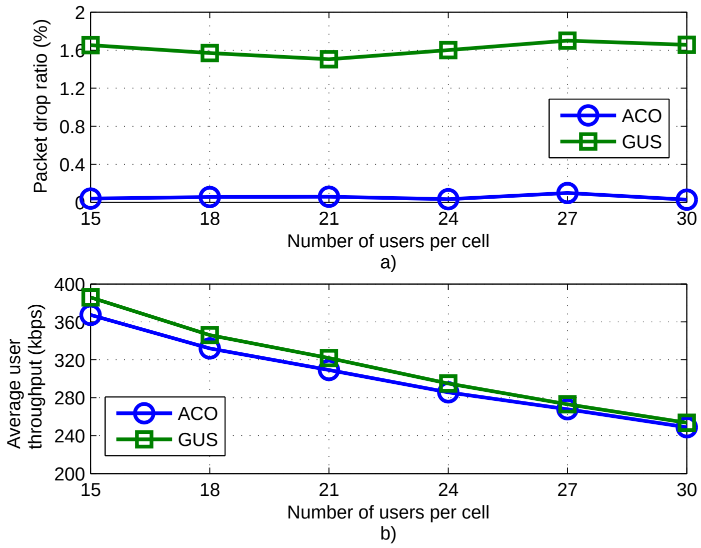


Figure 3.3: a) Cell-edge packet drop ratio of VoIP users b) Cell-edge average user throughput of BE users

3.5 Conclusion

In this paper, we focus on the downlink of multi-cluster CoMP OFDMA networks with multiple types of traffic patterns and diverse provisioning QoS requirements. A two-step joint scheduling scheme is proposed to address the utility-based combinatorial optimiza-

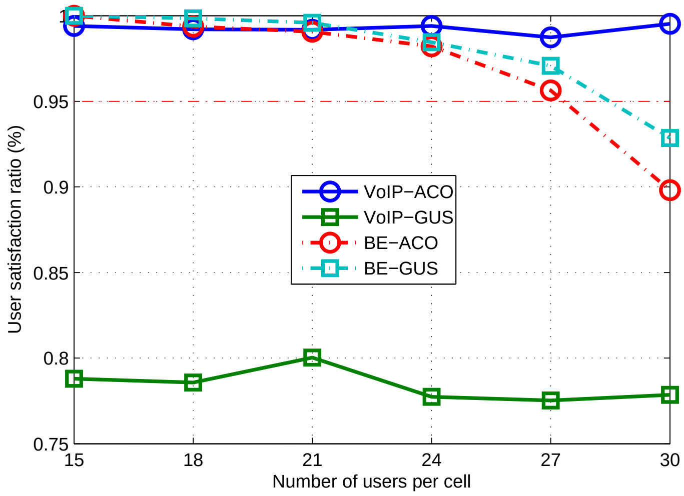


Figure 3.4: Cell-edge USR vs. cell user numbers

tion problem of joint scheduling in the network. By multi-cluster system level simulation, our proposed algorithm is shown to well approximates the optimal solution in terms of sum utility. Compared to the greedy user selection based algorithm, we demonstrate that the proposed joint scheduling algorithm yields higher sum utility, and achieves better performances in balancing average throughput of BE users and packet drop ratio of VoIP users at the cell-edge area with large number of users.

References

[1] 3GPP TR 36.814 V9.0.0, Further advancements for E-UTRA physical layer aspects, Mar. 2009.

[2] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tannol, Coordinated multi-point transmission/reception techniques for LTE-Advanced, *IEEE Wireless Commun. Mag.*, Jun. 2010.

[3] J. Li, H. Zhang, X. Xu, X. Tao, T. Svensson, C. Botella, and B. Liu, A novel frequency reuse scheme for coordinated multi-point transmission, *Proc. IEEE Veh. Technol. Conf.*, 2010.

[4] J. Li, X. Xu, X. Chen, X. Tao, H. Zhang, T. Svensson, and C. Botella, Downlink radio resource allocation for coordinated cellular OFDMA networks, *IEICE Trans. Commun*, vol. e93-b, 2010.

[5] H. Skjevling, D. Gesbert, and A. Hjørungnes, Low-complexity distributed multibase transmission and scheduling, *EURASIP J. Adv. Sig. Proc.*, 2008.

[6] R. Y. Chang, Z. Tao, J. Zhang, and C.-C. Jay Kuo, Multicell OFDMA downlink resource allocation using a graphic framework, *IEEE Trans. on Veh. Technol.*, vol. 58, Sep. 2009.

[7] B. Huang, J. Li, and T. Svensson, A Utility-Based Scheduling Approach for Multiple Services in Coordinated Multi-Point Networks, *IEEE Wireless Personal Multimedia Commun.*, 2011.

[8] M. Dorigo, and L. M. Gambardella, Ant colony system: a cooperative learning approach to the travelling salesman problem, *IEEE Trans. Evol. Comput.*, vol. 1, 1997.

[9] H. Ahmadi, Y. H. Chew, and C. C. Chai, Multicell multiuser OFDMA dynamic resource allocation using ant colony optimization, *IEEE Veh. Technol. Conf.*, Jul. 2011.

[10] M. Dorigo, T. Stutzle, Ant colony optimization, *MIT Press*, Cambridge, MA, 2004.

[11] Y. Zhao, X. Xu, Z. Hao, X. Tao, and P. Zhang, Resource Allocation in Multiuser OFDM System Based on Ant Colony Optimization, *IEEE Wireless Commun. & Networking Conf.*, Jul. 2010.

[12] K. I. Pedersen, T. E. Kolding, and P. Mogensen, QoS considerations for HSDPA and performance results for different services, *Proc. IEEE Veh. Technol. Conf.*, Sep. 2006.