Radio Resource Management in LTE-Advanced Systems With Carrier Aggregation

by

Ran Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In order to meet the ever-increasing demand for wireless broadband services from fast growing mobile users, the Long Term Evolution -Advanced (LTE-A) standard has been proposed to effectively improve the system capacity and the spectral efficiency for the fourth-generation (4G) wireless mobile communications. Many advanced techniques are incorporated in LTE-A systems to jointly ameliorate system performance, among which Carrier Aggregation (CA) is considered as one of the most promising improvements that has profound significance even in the upcoming 5G era. Component carriers (CCs) from various portions of the spectrum are logically concatenated to form a much larger virtual band, resulting in remarkable boosted system capacity and user data throughput.

However, the unique features of CA have posed many emerging challenges as well as span-new opportunities on the Radio Resource Management (RRM) in the LTE-A systems. First, although multi-CC transmission can bring higher throughput, it may incur more intensive interference for each CC and more power consumption for users. Thus the performance gain of CA under different conditions needs fully evaluating. Besides, as CA offers flexible CC selection and cross-CC load balancing and scheduling, enhanced RRM strategies should be designed to further optimize the overall resource utilization. In addition, CA enables the frequency reuse on a CC resolution, adding another dimension to inter-cell interference management in heterogeneous networks (HetNets). New interference management mechanisms should be designed to take the advantage of CA. Last but not least, CA empowers the LTE-A systems to aggregate the licensed spectrum with the unlicensed spectrum, thus offering a capacity surge. Yet how to balance the traffic between licensed and unlicensed spectrum and how to achieve a harmony coexistence with other unlicensed systems are still open issues.

To this end, the dissertation emphasizes on the new functionalities introduced by CA to optimize the RRM performance in LTE-A systems. The main objectives are four-fold: 1) to fully evaluate the benefits of CA from different perspectives under different conditions via both theoretical analysis and simulations; 2) to design cross-layer CC selection, packet scheduling and power control strategies to optimize the target performance; 3) to analytically

model the interference of HetNets with CA and propose dynamic interference mitigation strategies in a CA scenario; and 4) to investigate the impact of LTE transmissions on other unlicensed systems and develop enhanced RRM mechanisms for harmony coexistence.

To achieve these objectives, we first analyze the benefits of CA via investigating the user accommodation capabilities of the system in the downlink admission control process. The LTE-A users with CA capabilities and the legacy LTE users are considered. Analytical models are developed to derive the maximum number of users that can be admitted into the system given the user QoS requirements and traffic features. The results show that with only a slightly higher spectrum utilization, the system can admit as much as twice LTE-A users than LTE users when the user traffic is bursty. Second, we study the RRM in the single-tier LTE-A system and propose a cross-layer dynamic CC selection and power control strategy for uplink CA. Specifically, the uplink power offset effects caused by multi-CC transmission are considered. An estimation method for user bandwidth allocation is developed and a combinatorial optimization problem is formulated to improve the user throughput via maximizing the user power utilization. Third, we explore the interference management problem in multi-tier HetNets considering the CC-resolution frequency reuse. An analytical model is devised to capture the randomness behaviors of the femtocells exploiting the stochastic geometry theory. The interaction between the base stations of different tiers are formulated into a two-level Stackelberg game, and a backward induction method is exploited to obtain the Nash equilibrium. Last, we focus on the mechanism design for licensed and unlicensed spectrum aggregation. An LTE MAC protocol on unlicensed spectrum is developed considering the coexistence with the Wi-Fi systems. The protocol captures the asynchronous nature of Wi-Fi transmissions in time-slotted LTE frame structure and strike a tunable trade-off between LTE and Wi-Fi performance. Analytical analysis is also presented to reveal the essential relation among different parameters of the two systems.

In summary, the dissertation aims at fully evaluating the benefits of CA in different scenarios and making full use of the benefits to develop efficient and effective RRM strategies for better LTE-Advanced system performance.

Acknowledgement

There are many people I wish to express my sincere gratitude and appreciation who have made this thesis successful and meaningful. First of all, I wish to express my sincere gratitude to my supervisors, Professor Xuemin (Sherman) Shen and Professor Liang-liang Xie, for their invaluable and constant guidance throughout my graduate studies in University of Waterloo. During years of extensive training they provided me, they not only inspired me on how to conduct research in the cutting-edge field, but also profoundly shapes my thinking to be precise, logic, and sharp. More importantly, they recruited a smart, lovely and charming lady that I will cherish for my whole life as a husband.

I am grateful to have an examining committee with distinguished scholars. I would like to thank Professor Zhou Wang, Professor Fakhri Karray, and Professor Xinzhi Liu for their helpful comments and insightful questions. I would like to thank Professor Abdallah Shami from Western University for his commitment and serving as my external examining committee member.

I would like to thank all my friends and colleagues in the Broadband Communications Research Group, University of Waterloo. My sincere appreciation is extended to all friends and co-workers for their friendship and interaction. Special thanks go to Professor Lin X. Cai from Illinois Institute of Technology, Professor Rongxing Lu from Nanyang Technological University, and Professor Hao Liang from University of Alberta for their great encouragement and constructive discussions along the way. I am also grateful to Professor Yu Cheng from Illinois Institute of Technology, Dr. Zhongming Zheng, Dr. Ning Lu, Dr. Ning Zhang, Dr. Amila Gamage, Dr. Jian Qiao, Dr. Qinghua Shen, Qiang Ye, Nan Cheng, Kuan Zhang, Jianbing Ni, and Wenchao Xu for their expertise and friendship.

Last, but not the least, I am much obliged to my parents, Guangshui Zhang and Jinghong Zhang, my little sister, Yiwei Zhang, and my beloved wife, Dr. Miao Wang, for their unconditional love, understanding and support in my life.

To my parents, my little sister and my wife

Contents

List of Figures

List of Tables

Abbreviations

3GPP	the Third Generation Partnership Project
ACCS	Autonomous Component Carrier Selection
CA	Carrier Aggregation
CC	Component Carrier
CoMP	Coordinated Multi-Point Transmission
DRX	Discontinuous Reception
EC	Equivalent Capacity
eICIC	enhanced Inter-cell Interference Coordination
FBS	Femtocell Base Station
HARQ	Hybrid Automatic Repeat Request
HetNet	Heterogeneous Networks
IMT	International Mobile Telecommunication
ISP	Internet Service Provider
ITU	International Telecommunication Union
LBT	Listen-before-Talk Mechanism
LTE-A	Long Term Evolution - Advanced
LTE-U	Long-Term Evolution Unlicensed System
MAC	Medium Access Control
MBS	Macro Base Station

MIMO	Multi-Input-Multi-Output
OFDMA	Orthogonal Frequency Division Multiple Access
PCell	Primary Cell
PDCP	Packet Data Convergence Protocol
PHY	Physical Layer
(P)RB	(Physical) Resource Block
PSU	Partial Spectrum Usage
QoS	Quality of Service
RLC	Radio Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
SBS	Small Base Station
SCell	Secondary Cell
SC-FDMA	Single-Carrier Frequency Division Multiple Access
TTI	Transmission Time Interval
UE	User Equipment

Chapter 1

Introduction

With the challenges presented by the ever increasing use of "smart" wireless devices that require significantly higher spectral resources than conventional cell-phones, providing high quality of service for mobile applications in a cost-effective manner becomes increasingly important for operators to meet consumer needs. Towards this end, the International Telecommunication Union (ITU) has initiated a global standard initiative - International Mobile Telecommunication - Advanced (IMT-A) for 4G mobile communications in 2007 [1]. IMT-A systems include exciting new capabilities for providing a wide range of telecommunication services and applications, stressing improved quality of service (QoS) and worldwide development. To support enhanced user and service demands, peak data rate targets of 100 Mb/s for high mobility and 1 Gb/s for low mobility are established for IMT-A.

In response to ITUs invitation of candidate proposals for IMT-A, the Third Generation Partnership Project (3GPP) started Long Term Evolution -Advanced (LTE-A) in 2008 [2] which was ratified as an IMT-A technology in November 2010. LTE-A builds upon the 3GPP Release 8/9 specification, known as LTE [3], which is widely accepted as the transition standard from 3G to 4G. Although LTE systems could provide peak data rates of 300 Mb/s in the downlink (highest capability terminals) and 75 Mb/s in the uplink with scalable bandwidths of up to 20MHz, the the IMT-A data-rate requirements are still not satisfied. To fulfill the IMT-A requirements, LTE-Advanced shall support wider bandwidths than LTE

(e.g., up to 100 MHz). One approach is to simply extend the transmission bandwidth and use a single carrier. In practice, such a large portion of continuous spectrum is rarely available. Carrier aggregation (CA) [4–6], where multiple component carriers (CCs) of smaller bandwidth are aggregated, is an attractive alternative to increase data rate. Additional advantages are offered by CA in terms of spectrum efficiency, deployment flexibility, backward compatibility, and more. By aggregating non-contiguous carriers, fragmented spectrum can be more efficiently utilized. Various deployment scenarios for homogeneous and heterogeneous networks are supported by CA with proper utilization of different carriers. With each CC being LTE compatible, CA allows operators to migrate from LTE to LTE-A while continuing service to LTE users. Both implementation and specification efforts are minimized by reusing the LTE design on each of the CCs.

In this section, we will first present a brief overview of the LTE-A systems and the CA technology. The related works on Radio Resource Management (RRM) in LTE-A systems with CA are then introduced, in which the existing challenges and problems to be solved are proposed as part of our research motivations.

1.1 Overview of LTE-A Systems

1.1.1 Architecture of LTE-A Systems

Historically, mobile networks have consisted of a single homogeneous "layer" of macro cells of broadly uniform size. However, as the density of user equipments (UEs) increases, smaller cells are needed to deliver sufficient capacity. Since the small cells are introduced in a pre-existing network of macro cells, the result is a heterogeneous network (HetNet) of macro and small cells operating at different transmission powers and with different coverage areas. The HetNet architecture of the LTE-A systems are shown in Fig. 1.1.

The system layout of LTE-A is a HetNet where one macro cell is overlaid with a group of small cells and relays. A macrocell base station (MBS) in LTE-A systems is called the

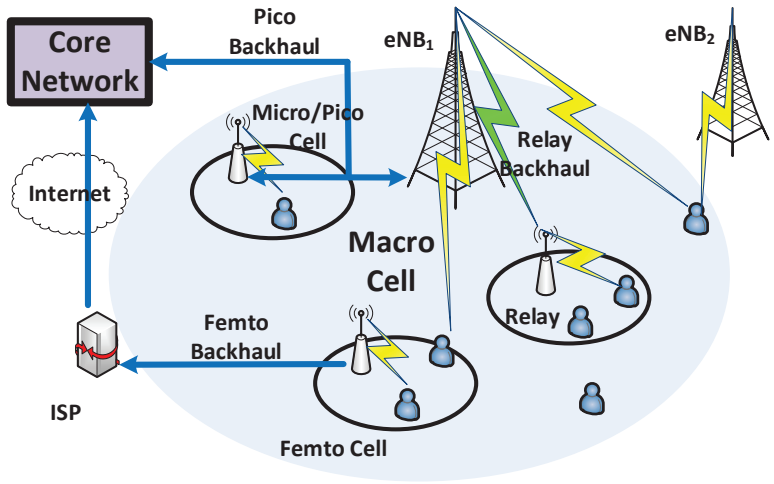


Figure 1.1. The Architecture of LTE-A Systems

enhanced NodeB (eNB) since the capability of the base station (BS) is enhanced with R-RM and part of the mobility management functionalities, which are usually performed in the radio network controller (RNC) in the 3G cellular radio access network [7]. An additional functionality is introduced in the LTE-A systems, which configures multiple MBSs to cooperatively communicate with one cell-edge user for performance improvement [8].

The small BSs (SBSs) operate on the same spectrum with the MBSs and are typically installed to extend the coverage to indoor areas where outdoor signals do not reach well or to add network capacity in areas with very dense service usage, e.g., train station. SBSs can provide high-data-rate services to the users in a small coverage range with relatively small transmission power, thus being low-cost and easy to be deployed. The small cells are basically divided into three types, i.e., the microcells, the picocells and the femtocells. The microcells and picocells are usually deployed by the vendors and has direct backhaul to the eNBs (MBSs) through wireline, which means the RRM and mobility management of the picocells are controlled directly by the MBSs. While the femtocell is usually deployed by the end users and its backhaul to the cellular core network is through the Internet service providers (ISPs). Thus the femtocells have larger signaling latency and the deployment is harder to be predicted and controlled by the MBSs. Generally, the transmission range of the

microcells is less than $2km$. A picocell, on the other hand, is $200m$ or less, and a femtocell is on the order of $10m$ [9]. Besides, relays are deployed to improve the performance of cell-edge users. Different from the small cells, the backhaul link of the relays is wireless, meaning that all the control signalings and data flows between relays and the eNB must be transmitted wirelessly, thus occupying some bandwidth resources of the eNB. However, compared to the small cells, relays have the lowest deployment cost.

1.1.2 Characteristics of LTE-A Systems

Fulfilling the standard requirements of the IMT-A, LTE-A systems can achieve much higher system capacity and spectrum efficiency than the LTE systems. Table 1.1 gives a comprehensive technical comparisons among LTE, LTE-A and the IMT-A requirements [10].

Items	**LTE**	**LTE-A**	IMT-A
Downlink peak rate with CA	300 Mb/s	3 Gb/s	1 Gb/s
Uplink peak rate with MIMO	75 Mb/s	500 Mb/s	N/A
Supported bandwidth (MHz)	Up to 20	Up to 100	5-20
Radio access technology	Dowlink: OFDMA Uplink: SC-FDMA	Downlink: OFDMA Uplink: hybrid OFDMA/SC-FDMA	N/A
Downlink spectrum efficiency (bps/Hz)	16.3	30.6	15
Uplink spectrum efficiency (bps/hz)	4.3	16.8	6.75

Table 1.1 Technical comparison between LTE and LTE-A with IMT-A requirements

From the table, it can be seen that LTE-A can achieve much higher peak data rate and spectrum efficiency than LTE and satisfy all the IMT-A requirements. The LTE system supports scalable bandwidths of up to 20 MHz providing peak data rates of 300 Mb/s in the downlink (highest capability terminals) and 75 Mb/s in the uplink. However, the LTE systems still cannot satisfy the IMT-Advanced requirements in peak data rates and spectral efficiency. To settle this problem, LTE-A exploits CA to combine at most five CCs (i.e., LTE carriers), each has a bandwidth of up to 20MHz [11], to aggregate small bandwidth segments into a wider bandwidth for one transmission, thus enhancing the system capacity by multiple times. At the same time, more advanced MIMO techniques are adopted in the

system with more spatial layers (concurrent transmitted information flows) in both downlink and uplink [10]. Together with the MIMO technique with full spatial multiplexing gain, the peak data rate of LTE-A can achieve 3Gb/s in downlink and 500 Mb/s in uplink [12].

For the radio access technologies, both LTE-A and LTE adopt the orthogonal frequency division multiple access (OFDMA) [13] in the downlink; however, in the uplink, LTE adopts the single carrier frequency division multiple access (SC-FDMA) [13], while LTE-A adopts hybrid OFDMA/SC-FDMA. The different in the uplink is due to the backward compatibility in the design of LTE-A systems. Building upon the LTE systems, a LTE-A system must be designed in such a way that the legacy LTE users who can only use SC-FDMA in the uplink can operate normally in the system, being compatible with all the additional enhanced functionalities.

1.2 Carrier Aggregation (CA)

As one of the most momentous techniques in LTE-A, Carrier Aggregation (CA) allows scalable bandwidth extension via aggregating multiple smaller band segments, each called a Component Carrier (CC), into a wider virtual frequency band to transmit at higher rates [14]. The following subsections give a brief introduction on the CA design principles and management characteristics.

1.2.1 Design Principles

The design of 3GPP LTE-A CA considers various aspects including backward compatibility, system modification, implementation complexity, and so on with the following design principles [14].

1. **Backward Compatibility** - Backward compatibility is critical for LTE-A CA to migrate smoothly from LTE and reuse the LTE design to the most extent. Each CC in LTE-A is LTE backward compatible, i.e, accessible by the LTE UE. The complete set

of LTE downlink transmissions are performed on each CC following the LTE physical procedure and specifications.

2. **Minimum Protocol Modifications** - From the aspect of user-plane protocols, the CCs are invisible to the Packet Data Convergence Protocol (PDCP) and radio link control (RLC) layers. The multiple CCs are only different data transmission pipes managed by a single scheduling entity at the medium access control (MAC) layer. Each CC has its own LTE-compatible hybrid automatic repeat request (HARQ) processes for the physical (PHY) layer transmissions. The PHY and MAC design for 3GPP LTE-A supports up to 5 CCs despite of the CA types.

3. **Limited Control Procedure Impact** - In the control-plane aspect, radio resource control (RRC) entity assigns the radio management information from the network to the UE. At a given time instance, one UE is in either RRC_IDLE or RRC_CONNECTED state. One UE can transmit/receive data to/from the network only when it is RRC_CONNECTED. One RRC_IDLE UE shall transit to RRC_CONNECTED state by establishing an RRC connection following the LTE procedure before being able to transmit on multiple CCs. Hence, LTE-A CA does not change the RRC_IDLE procedures; nor does it impact the establishment procedure of an RRC connection.

1.2.2 Characteristics of CA

Types of CA

LTE-A systems support three types of CA: intra-band contiguous CA, intra-band non-contiguous CA and inter-band CA, as shown in Fig. 1.2. For the first type, all the aggregated CCs are located within the same frequency band (e.g., the GSM @900MHz) and are contiguous one by one. The CCs in the second type are also located within the same frequency band but may not be contiguous to each other. As the CCs of the first two types are both located within the same band, the radio characteristics (e.g., the channel fading statistics) of each CC can

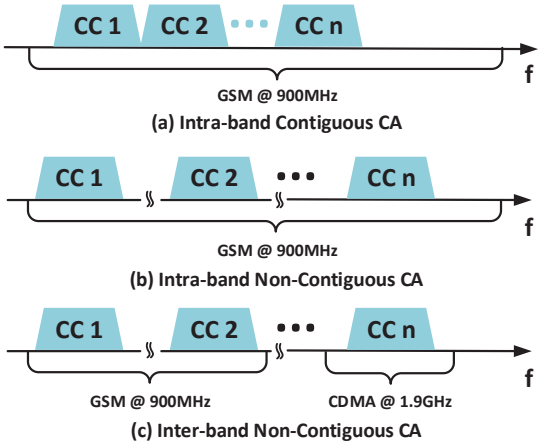


Figure 1.2. Three different CA types.

be considered identical. In the third type, CCs can be located in different frequency bands (e.g., GSM and CDMA bands), thus having different radio characteristics which should be carefully considered into the RRM framework for inter-band CA.

Adaptive CC Selection/Configuration

In LTE-A CA, a CC is often referred to as a serving cell and is treated as such by the higher layer procedures. For frequency-division duplex (FDD), a serving cell comprises a pair of different carrier frequencies for downlink and uplink transmissions. For time-division duplex (TDD), a serving cell is defined for a single carrier frequency where downlink and uplink transmissions occur in different transmission time intervals (TTI). Each UE has a single serving cell that provides all necessary control information and fuctions, such as mobility and security information, RRC connection maintenance, etc. This serving cell is referred to as the primary cell (PCell). Other serving cells are referred to as secondary cells (SCells).

Cell management is the control procedure in layer 3 (i.e., the network layer) enabling the network to add/remove/change an SCell or to switch the PCell of UE. AN RRC_IDLE UE establishes an RRC connection toward a serving cell, which automatically becomes its PCell. Depending on the carrier where initial access is performed, different UEs in a CA system may have different PCells. With the RRC connection on the PCell, the network can

further configure one or more SCells for UE within the UE CA capability to meet traffic demands. The necessary information, including system information, of an SCell is conveyed to the UE via dedicated RRC signaling. Addition, removal, and reconfiguration of SCells to a UE are also performed via dedicated RRC signaling. The network can further change the PCell of a UE, for example, to improve the link quality of the PCell on which critical control information is sent or to provide load balancing among different SCells. PCell change in CA can only be performed via the handover procedure. PCell change does not necessarily require UE to switch to single-CC operation. Intra-LTE handover in LTE-A allows the target PCell to configure one or more SCells for UE to use immediately after handover.

Dynamic CC Activation/Deactivation

Cell activation/deactivation is a mechanism in MAC layer aiming to reduce UE power consumption in CA on top of discontinuous reception (DRX) [15], which is already supported in LTE Release 8/9. DRX puts UE into power saving mode when the UE is not expected to receive data from the network. According to network configuration and ongoing HARQ processes, UE determines the DRX ON/OFF duration common to all serving cells. To further reduce UE battery consumption, an SCell in CA can be activated or deactivated [16]. For a deactivated SCell, UE does not receive any downlink signal; nor does the UE transmit any uplink signal. Conversely, for an activated SCell, UE performs normal activities for downlink reception and uplink transmission. The SCell activation/deactivation is enabled by a combination of explicit and implicit means where the network can issue an activation/deactivation command in the form of a MAC control element (CE), or the UE autonomously deactivates a serving cell upon timer expiry. Serving cell activation/deactivation is performed independently for each SCell, allowing UE to be activated only on a necessary set of SCells. Activation/deactivation is not applicable for the PCell since the functions provided by the PCell require it to always remain activated when the UE has an RRC connection to the network.

1.2.3 Spectrum Access in LTE-A Systems with CA

Spectrum Access on CC Resolution

Each BS can dynamically select from a finite set of available CCs. Each MBSs are accessible to all the available CCs while for the densely deployed SBSs, each node only uses a subset of available CCs. This is the best configuration for optimizing the system performance as there is severe interference coupling between those nodes. Notice that by conducting the adaptive frequency reuse on CC resolution, both data and control channels experience benefits within a single CC.

Spectrum Access on PRB Resolution

The bandwidth structure of one single CC is shown in Fig. 1.3. As shown in Fig. 1.3, the

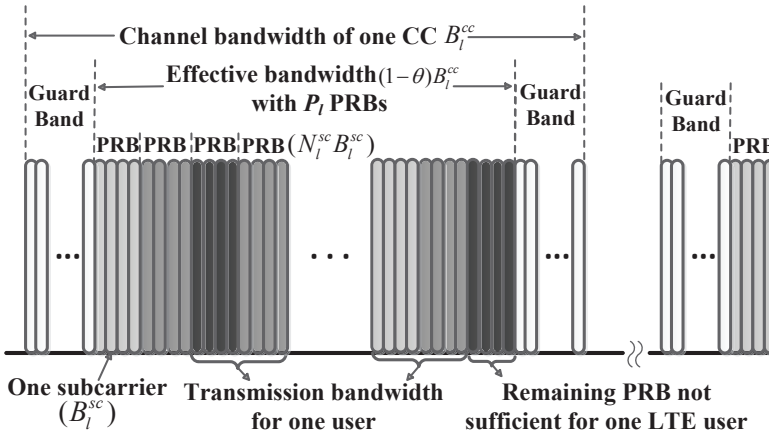


Figure 1.3. Bandwidth structure for LTE-A systems based on OFDMA.

channel bandwidth B_l^{cc} of the lth CC is up to 20MHz and contains two parts: the guard bands (GBs) and the transmission bandwidth. As specified in [17], the GBs are set on both sides of each CC to avoid interference caused by Doppler Shift and Frequency Aliasing Effect in real systems [18]. No effective data will be transmitted in the GBs. The total percentage of GBs is denoted as θ. For the lth CC, the transmission bandwidth is divided into P_l PRBs, each composed of N_l^{sc} continuous subcarriers with bandwidth B_l^{sc}. The relationship of all

the above variables is given as

$$P_l = \frac{(1-\theta)B_l^{cc}}{B_l^{sc}N_l^{sc}},\ l = 1, 2,\ ...,\ n, \tag{1.1}$$

where n denotes the number of CCs. The assignment of PRBs for each user is determined according to the throughput requirement and channel conditions. The interference from other users in the same cell is ignorable due to the orthogonality of PRBs. For LTE users, the transmission bandwidth may not be fully utilized as the remaining PRBs in one CC may not be sufficient to serve any more users. However, for LTE-A users, these unused PRBs in different CCs could be combined together to jointly serve a user through the CA technique. In this way, the LTE-A users could achieve higher spectrum usage than the LTE users. In this work, such PRBs are referred to as *semi-usage PRBs*.

1.3 RRM Framework for LTE-A Systems with CA

The RRM framework in LTE-A systems within our research scope involves the protocols and functionalities from layer 3 to layer 1. The overview of the hierarchical user plane and the corresponding mapping of the most essential RRM functionalities are shown in Fig. 1.4.

1.3.1 Hierarchical User Plane

The left part of Fig. 1.4 shows the user plane of CA within our research scope. Each user has at least one radio bear, carrying one data flow for the user. Different radio bears have different QoS provisions for service differentiation. Each radio bearer is associated with one data packet convergence protocol (PDCP) and radio link control (RLC), which are inherited from LTE Release 8 [19]. These two protocols perform functionalities such as robust header compression (ROHC), security, segmentation, and outer automatic repeat request (ARQ). The interface between RLC and the medium access control (MAC) is refered to as *logical*

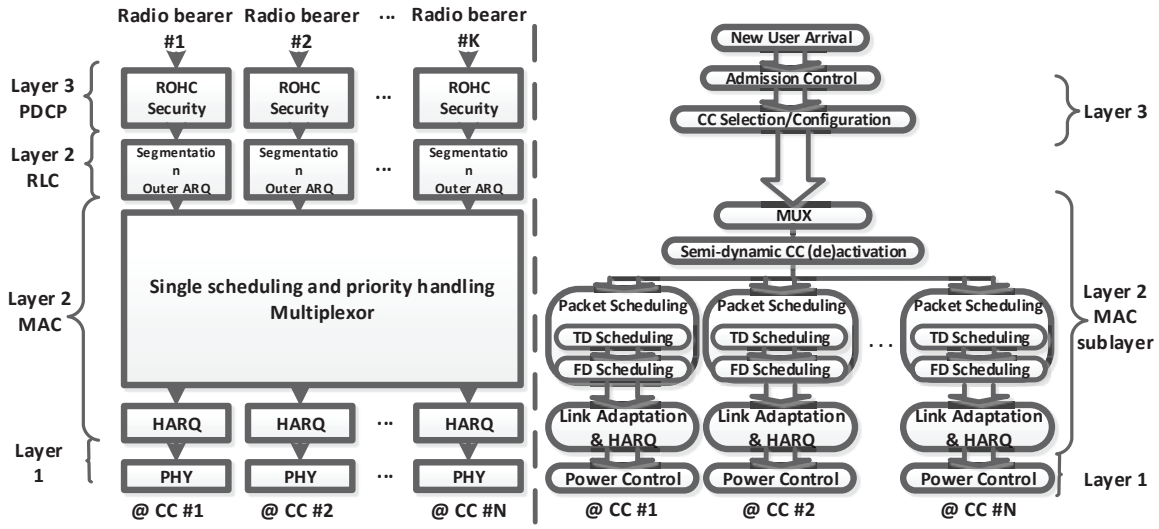


Figure 1.4. Hierarchical user plane and the corresponding RRM functionalities

channels, which are further divided into control channels and traffic channels. Each user has one MAC entity, which multiplexes (MUX) the data from all the logical channels to the user (or eNB), and distributes the data to transmissions on different available CCs. As illustrated in Fig. 1.4, each CC has a separate HARQ entity, essentially meaning that the transmission and possible retransmissions of one packet are on the same CC. The interface between MAC and the physical layer (PHY) is denoted as *transport channels*, each mapping to one or more logic channels. The transport channels are also separate for each CC. Data transmissions on different CCs can adopt independent modulation and coding schemes (MCSs). As a result, independent link adaptation (LA) per CC is enabled to benefit from optimally adjusting transmissions on different CCs according to the corresponding channel conditions. The power control for different CCs can also be independent, making it possible for one eNB to have diverse coverage levels on different CCs [14].

For the control-plane protocol of LTE-A CA, similarly as LTE Release 8, each user has one radio resource control (RRC) entity, which is independent of the number of CCs [19]. To fulfill the function of CA, a few more functionalities have been added into RRC, which will be elaborated in the next subsection.

1.3.2 RRM Functionalities and Considerations

The main RRM functionalities of LTE-A are presented in the right part of Fig. 1.4. The LTE-A RRM framework has many similarities with that of LTE Release 8/9. When a user arrives, the admission control is performed at the eNB before the establishment of the RRC connection. New radio bearers are then created and the corresponding QoS parameters are configured. Specifically, a new functionality is introduced into LTE-A CA, which is referred to as CC selection or configuration in the following. CC selection configures a set of CCs for each user to be scheduled on. This function is an important tache to optimize system performance over the entire CC set, as well as controlling the power consumption. The CC selection takes as inputs the information of QoS parameters, UE capabilities, radio bearer configuration, CC load and channel conditions. Among the inputs, the CC load conditions are used for cross-CC load balancing [19], and the QoS parameters specify the user service types (e.g., best-effort and guaranteed-bit-rate services) and the QoS requirements. For a UE, assigning more CCs can significantly increase its throughput, but resulting in higher UE power consumption and signaling overhead, as well as more interference to other UEs operating on the assigned CCs. Therefore, potential freedom is left for researchers to design optimal CC selection algorithms to achieve various performance objectives.

Same as LTE Release 8, in MAC layer of LTE-A systems, all the data flows from different users on one CC are multiplexed together, and the packet scheduling (PS) entity in each CC will allocate PRBs to the attached users in every Transmission Time Interval (TTI) of $1ms$. Basically, the PS takes advantage of multi-user frequency-domain scheduling diversity by preferentially allocating PRBs to users that are perceiving good channels. The PS process consists of two phases, i.e., time-domain PS and frequency-domain PS. The entity first determines which subset of users should be scheduled in the next TTI and then determines which part of and how much bandwidth should be allocated to each scheduled user.

However, there are two main differences between LTE-A PS and LTE PS. First, the PS in LTE-A is allowed to schedule users across multiple CCs. The scheduling could be

done in parallel in different CCs, but with some coordination to ensure fairness and joint control for users scheduled on multiple CCs [20]. This cross-CC scheduling functionality offers higher flexibility and better overall system performance for transmissions of control and data information across multiple CCs. Second, on top of the regular PS in each CC, an additional functionality is designed to dynamically (de)activate CCs configured as SCells for different users, which is controlled by MAC signalings through the PCell [16]. A user is only schedulable on its configured and activated SCCs and does not report channel state information (CSI) for link adaptation and frequency-domain PS through the SCCs. With this functionality, the number of CCs that a user can be scheduled on can be dynamically adjusted in tens of ms according to the instantaneous cell load conditions and QoS requirements. In this fashion, the UE power can be further preserved.

In addition, within each CC, link adaptation is performed to dynamically adjust the UE's modulation and coding schemes according to the time-varying channel conditions; HARQ is performed to manage packet retransmissions when packet loss or error occurs. At last, the power control entity in Layer 1 will decide the UE transmission power on each CC either independently or coordinately.

1.4 Organizations and Contributions

The remainder of the dissertation is organized as follows: Chapter 2 introduces the research topics on RRM in LTE-A systems with CA, and elaborates some related works to our most interest for each topic. Chapter 3 presents our research results on demonstrating the advantages of CA in improving the system limiting capabilities, i.e., the system user-accommodation capabilities. Chapter 4 investigates resource (i.e., spectrum and power) optimization in single-tier LTE-A CA-based systems. Chapter 5 studies interference management problems in multi-tier LTE-A systems with CA. The conclusions and future research directions are presented in Chapter 7. Finally, Chapter 8 concludes the dissertation with our related publications.

Chapter 2

Research Topics and Related Works on RRM in LTE-A Systems with CA

Many research issues emerge in the realization process of CA-based LTE-A systems, among which RRM is very imperative in providing guidelines to fully utilize the network resources. Since OFDMA/SC-FDMA is adopted as the access technology in LTE-A standard, RRM in LTE-A systems can date back to the studies on RRM in OFDMA systems, where subcarrier is the minimum bandwidth allocation resolution. Most of the related studies [21–23] mainly focus on reorganizing the limited network resources to optimize the network performance. For instance, in [21], a novel scheme for the allocation of subcarriers, rates, and power was proposed to maximize the aggregated data rates. In [22], the energy efficiency problem was investigated for cognitive radio systems under the QoS constraints. In [23], the uplink relay selection problem was discussed to enhance the total achievable throughput under total power constraint. We refer to [24–26] as comprehensive surveys on RRM in OFDMA systems.

However, these works can not be directly applied to LTE and LTE-A systems due to the unique features of LTE-based systems. Different from subcarriers in OFDMA systems, the minimum resolution in LTE-based networks is Physical Resource Block (PRB) which is composed of 12 consecutive OFDM subcarriers. One user could be assigned with several PRBs but all the subcarriers in one PRB must be assigned to the same user. This feature

significantly reduces the control overhead and meanwhile improves the spectrum utilization. It also gives rise to new challenges upon the design of scheduling and power allocation schemes for both control channels and data channels. Thus, the flexibility and efficiency of the existing works should be re-evaluated, and new schemes particularly for PRB allocation and power control should be considered. For instance, since the minimum bandwidth allocation unit in LTE is the PRB, [27] put forward a distributed and coordinated PRB and power allocation scheme to mitigate the intercell interference in LTE. As the control channel structure is updated in LTE over OFDMA systems, [28] showed different conditions when an LTE system is data-channel limited or control-channel limited. [29] further gave a comprehensive overview of downlink RRM for LTE systems.

Compared with LTE networks, the adopted CA technique in LTE-A systems also brings new challenges and favorable opportunities for RRM. With CA, an LTE-A user could operate on multiple CCs concurrently to have larger throughput. But operating on more CCs means more energy consumption at the user terminal and more intensive interference to other users operating on the same CCs. Thus, how to dynamically decide which CCs should be assigned to different users is a critical issue given the tradeoff among throughput, energy consumption and interference intensity. Besides, cross-CC load balancing and scheduling should also be deliberated to achieve better overall network resource utilization.

In light of these challenges, this chapter intends to uncover the main imperative issues on RRM of CA-based LTE-A systems in a systematic manner, and outline our research logic path as a whole towards tackling the open challenges.

2.1 Benefit Demonstration of CA

In CA-based LTE-A scenarios, the legacy LTE users and LTE-A users could co-exist and share the system bandwidth up to 100MHz, where LTE users can only be scheduled on one CC at any time while LTE-A users can transmit on multiple CCs concurrently with larger possible throughput. Then there rises a problem: given this advantage, how much the

performance of LTE-A users can surpass that of LTE users in different scenarios should be carefully studied to demonstrate the benefits of CA.

This problem can be explored from two perspectives, i.e., the user-centric perspective and the system-centric perspective. On one hand, given a set of users or an arrival-departure process of users, the user-centric performance (e.g., average throughput, user service delay [30] and power efficiency [31] in the user terminal) can be compared between LTE users and LTE-A users under different cell load conditions. There have already been many works following this thought [19, 20, 32]. For example, the work [19] focuses on comparing the throughput performance via simulations between the legacy LTE users and LTE-A users in a multi-CC LTE-A system. Simulation results shows that LTE-A users outperform the LTE users in terms of the average throughput and cell-edge throughput, especially in light cell load conditions. The work suggests that the number of CCs accessible per LTE-user should vary with the cell load conditions for better performance. In [20], a joint carrier load balancing and packet scheduling scheme is put forward, comparing the average user throughput when LTE users and LTE-A users coexist under cross-CC scheduling. It is shown that the LTE-A users can achieve significantly higher throughput under different conditions when coexisting with LTE users. In [32], leveraging the stochastic geometry theory, analytical analysis is developed to derive the user SINR distributions and average throughput for LTE and LTE-A users, respectively. The results indicate that although LTE-A users can achieve higher throughput than LTE users, the perceived SINR is worse in a single CC for LTE-A users due to multi-CC transmissions. Thus, the performance gain may become smaller with larger number of cell users as the SINR situations for LTE-A users become worse faster.

On the other hand, this issue can be investigated from the system-centric perspective, i.e., to study the system limiting capabilities. For example, in the admission control process, given the same system settings and user QoS requirements, the user accommodation capabilities of LTE-A systems can be analyzed to show how many LTE users or LTE-A users can be admitted into the system. However, few works have dabbled in this issue from this perspective whereas it can provide valuable guidelines for optimizing resource utilization

and guaranteeing QoS provisioning at the same time. To this end, one of our research goals is to explore the performance improvement in the system limiting capabilities in different scenarios when CA is adopted, while jointly considering the QoS requirements (e.g., loss probability, throughput requirement) and traffic descriptors (e.g., active probabilities) of different user types. Our research works [33, 34] have focused on this issue, where the LTE-A downlink admission control process is studied and a QoS-aware closed-form admissible region for heterogeneous user classes are derived (See Chapter 3).

2.2 CA in Single-Tier LTE-A Systems

For LTE-A CA, it is not good to always assign multiple CCs to every user. On the contrary, in some cases the multi-CC transmission can significantly counterbalance the performance gain. For instance, for a best-effort user, assigning as many CCs as possible may be less optimal when the cell load becomes quite heavy, since the extremely crowded situation in one CC can considerably reduce the average per-CC user throughput, resulting in a lower aggregate user throughput [19] [35]. On the other hand, for a user with a guaranteed throughput requirement, when the cell load becomes light, only assigning one CC is enough to satisfy its minimum throughput requirement and assigning multiple CCs only consumes more UE power in vain [35]. Therefore, it is essential to assign a proper CC subset dynamically and adaptively according to the instantaneous cell load and user service types. Besides, for layer-2 packet scheduling (PS), independent PS per CC will lead to unbalanced user performance between the legacy LTE users and LTE-A users since a scheduler allocates the resources only based on the knowledge of its own CC. To achieve a certain level of fairness, cross-layer PS is desirable by jointly considering the resource allocation (RA) conditions from all the assigned CCs for a user.

There have been abundant research works related to dynamic CC management and cross-CC packet scheduling in both downlink [20, 36–38] and uplink [39–41], where extensive theoretical analysis and experimental results show that adaptive CC selection and cross-

CC packet scheduling can significantly enhance not only the system capabilities such as power efficiency and interference mitigation, but also the user performance in throughput and fairness. We refer to [42,43] as comprehensive overviews.

However, for the dynamic CC management, most of the existing literatures only put emphasize on the layer-3 adaptive CC selection/configuration function, and two important factors are usually less discussed or even neglected. The first one is the layer-2 dynamic CC (de)activation function. As aforementioned in Subsection 1.2.2, this function is performed in TTI level via MAC signaling to dynamically decide whether a UE should sleep on a particular assigned CC, i.e. stop monitoring the Physical Downlink Control Channel (PDCCH) with no transceiving. Some pioneering ideas have already been proposed in industry [44–46], which principally design the implementation methods and signaling procedure. There are also a few works from academia [47,48], which discussed the impact of (de)activation periods and frequency on the energy saving efficiency and proposed CC-specific DRX mechanism considering multiple services. However, comprehensive performance analysis and evaluation are still lacked. Thus, realizing the great potential of this mechanism in significantly reducing the UE power consumption especially in the uplink, cross-layer dynamic CC management strategies should be designed to combine the layer-2 CC (de)activation mechanism with the layer-3 adaptive CC selection function. In this manner, higher power efficiency in the downlink and lower power consumption in the uplink can be achieved.

The second neglected factor is the power offset effects. Basically, when a UE is transmitting on multiple CCs simultaneously, the increased Peak-to-Average Power Ratio (RARR) and the inter-modulation [49] will lead to additional power consumption for the UE in the uplink, resulting in a non-neglectable reduction in UE's maximum transmission power. These effects, referred to as power offset effects in this dissertation, degrade the user performance inevitably. Thus, it is essential and challenging to consider the power offset effects into the dynamic CC management framework, especially for uplink CA to improve the UE power utilization. One recent work [50] has incorporated this effect into analysis and modeled the resultant power backoff as a constant. A threshold-based CC selection strategy was proposed

to improve the user throughput based on the user path loss and the number of available CCs in the cell. But the work does not consider how the time-variabilities of the offset effects impact the RRM performance in uplink CA. In fact, the user power offset should vary with the number of CCs, the number of instantaneous occupied PRBs in each TTI and the dispersion degree of the PRBs [49]. To this end, our research work [35] has proposed a joint uplink CC selection and power control scheme, which improves the average user throughput by maximizing the user power utilization with considering the time-variabilities (See Chapter 4).

2.3 CA in Multi-Tier LTE-A Systems

LTE-A systems support the coexistence of macro cells with small cells sharing the same frequency band, forming into multi-tier LTE-A systems, termed as HetNets in 3GPP terminology [10]. In HetNets, the macro-cells are accessible to all the users and deliberately planned to reduce the inter-cell interference as much as possible, while the small cells spread out irregularly mainly serving users within a small coverage. Providing high spatial reuse via cell splitting, such deployment can achieve substantial gains in coverage and capacity compared to the single-tier macro-only networks. However, since the small cells are deployed in an unplanned way, co-channel interference between macro and small cells (cross-tier interference) and among small cells themselves (co-tier interference) becomes a problem with higher magnitude and variability. Therefore, efficient interference coordination schemes must be carefully designed to realize the potential gains.

Currently, a majority of the existing works deal with this issue on the aspect of co-channel interference mitigation, where macro and small cells have access to all the bandwidth [51–53]. In [51], a distributed femtocell management architecture based on OFDMA is proposed to make a tradeoff between macrocell and femtocell capacities. [52] studies the outage probability of downlink femto-macro networks with a 3D-poisson model of random spatial distribution in an LTE environment. Further in [53], authors propose a novel joint cell

association and interference management schemes for LTE-A HetNet to maximize the sum utility of average rates while satisfying the users QoS. All the above works have to involve time-domain scheduling coordination between macro and small cells which has relatively high complexity.

Recently, the CA technology has been introduced to play a key role in inter-cell interference coordination by dynamically configuring different subset of CCs to different small cell base stations (SBSs) to realize interference avoidance instead of mitigation. Such spectrum management of small cells based on CA is also referred to as partial spectrum usage (PSU) on the CC resolution. As a totally new feature adopted in LTE-A systems, CA has the nature to enable simple, yet effective frequency domain interference management schemes for both data and control channels [54].

The interference management schemes in HetNets with CA mainly fall into three categories: centralized control management [55, 56], semi-autonomous control [57, 58] and the distributed autonomous control [59–62]. For centralized control management, the eNB is in charge of making RA decisions for the small cells jointly considering the cross-tier and co-tier interference as well as the QoS requirements of different user services. As the eNB needs to know the RRM-related knowledge of small cells, the centralized control strategies have strict requirements on the latency and capacity of the small cell backhaul connections for timely message exchange. For the semi-autonomous control strategies, the eNB does not directly control the RAs of small cells, but instead influence the RRM decisions in an indirectly manner. For instance, in [57], Duan *et al.* designed a game-theory-based price control strategy where MBSs influence the FBS behaviors by determining the user service prices for macrocells (MCells) and Femtocells (FCells). In [58], Bu *et al.* proposed to set an interference price for MBSs over FBSs based on the interference from FBSs; the FBSs then consider the price in its own RA to maximize their own utilities. Through such interaction, the interference between MCells and FCells can be effectively coordinated. For the distributed autonomous control strategies, the small cells do the RA merely based on the local obtained information, i.e., the cell load conditions, the perceived interference level, the

user service requirements, the available network resources, etc. The most typical strategy of such category is the autonomous CC selection (ACCS) [59, 60], which is inherently a fully distributed and dynamic interference management concept in CC resolutions. ACCS relies on the sensed interference levels and adding an additional CC must satisfy the premise that it will not influence the transmissions in neighboring cells to a certain extent. Many decentralized dynamic spectrum sharing and cooperation approaches can find their valuable roles in LTE-A HetNets with CA, e.g., game theory and reinforcement learning [63] [64].

Comparing the above three categories, the first one can achieve the best overall network performance, yet has the most strict requirements on the backhaul links, which are difficult to be satisfied in some occasions. The last category has the largest flexibility and least implementation/computation complexity, but may very likely lead to sub-optimal network performance. The most applicable scenarios for the third category include the indoor hotspots environments where the signal strength of small cells dominates that of the MCells, and the rural regions where MCells are not available and small cells are deployed in an ad hoc way. The semi-autonomous strategies strike a good tradeoff between the implementation complexity and network performance, but need advanced mathematical tools to model the interaction between MCells and small cells and determine the best equilibrium.

Although a multitude of works have been done, the interference management in HetNets with CA still needs more exploration. Up to now, most related works typically consider a static HetNet deployment where the locations and types of base stations are pre-defined. The carrier selection strategies are only dynamically adaptive to the traffic variabilities. Unfortunately, this cannot be always true in reality since the small cells like femtocells may appear and disappear at anytime and anywhere due to human activities or the instability of the power sources. The locational and and temporal randomness of the femtocells will bring considerable challenge to the robustness of the existing strategies. To investigate this problem, our research work [65] exploit a stochastic mathematical tool - stochastic geometry [66] [67] and the related concepts, to characterize the randomness of the co-deployment topology. Based on the stochastic topology model, a semi-autonomous interference management strategy is

developed to model the interplay between the MCells and FCells using game theory techniques in the context of PSU. By analytical analysis and comprehensive simulations, the work aims to provide significative insights on how the FCell randomness impact the interference management performance of HetNets with CA and show how to maximize the utilities of both MCells and FCells under different situations (See Chapter 5).

2.4 CA in Aggregating Licensed and Unlicensed Spectrum

With the proliferation of mobile devices and diverse mobile applications, wireless operators are experiencing phenomenal mobile data growth around the world. It is expected that by the year 2020, the industry need to be prepared for as much as 1000 times mobile traffic as the year 2010 [68]. In the upcoming 5G era, both the industry and academia are on the hunt for advanced solutions to boost the network capacity. The current available bandwidth in the licensed spectrum cannot afford the explosive mobile data demand, and excavating more capacity from other spectrum becomes indispensable. Thanks to the technique of CA, the reach of cellular systems could be extended from licensed-only operations to the unlicensed spectrum which can provide much more bandwidth. The technology which aggregates the licensed spectrum with the unlicensed spectrum for RRM leveraging the CA technique is named as LTE-Unlicensed (LTE-U) technology. An illustration of LTE-U rationales is showed in Fig. 2.1.

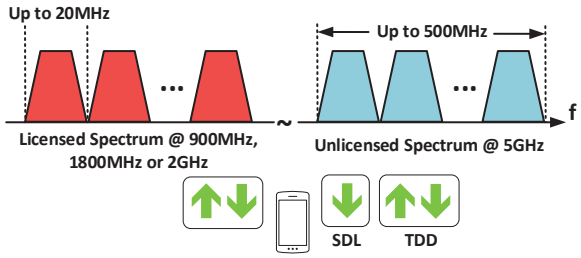


Figure 2.1. Deployment scenarios of LTE-U technology.

Due to the transmission power limitations [69] in unlicensed spectrum, the LTE-U tech-

nology is more suitable for a small area. Hence, the deployment of most interest is operator-deployed small cell which provides access to both licensed and unlicensed spectrum for indoor environment or outdoor hotspots. The aggregation of licensed and unlicensed spectrum can provide the small cell users with high-speed and seamless broadband multimedia services. During transmission, a licensed carrier, serving as Primary Component Carrier (PCC), and several unlicensed carriers, serving as Secondary Component Carriers (SCCs), are accessible to one user at one time. According to the user traffic demand and cell load, configuration information can be conveyed via PCC to dynamically remove/add SCCs.

There are two operation modes for LTE-U [70]: supplemental downlink (SDL) and time division duplex (TDD), as shown in Fig. 2.1. SDL mode is the simplest form of LTE-U where the unlicensed spectrum is only used for downlink data transmission since downlink traffic is typically much heavier than uplink traffic. In this mode, LTE eNB can perform most of the required operations for reliable communications, including detecting the unlicensed channel occupancy. In TDD mode, the unlicensed spectrum is used for both downlink and uplink, just like the LTE TDD system in licensed bands. TDD mode offers the flexibility to adjust the resources allocation between downlink and uplink, at the cost of extra implementation complexity on the user side.

Transmission on unlicensed spectrum is unstable since the "unlicensed" nature makes it hard for provisioning guaranteed QoS. To ensure the QoS and improve the user experience, the use of unlicensed spectrum in LTE-U must come with the use of licensed spectrum. With CA, the control-plane messages including radio resource control signallings and Layer-1 signallings, are always transmitted on the licensed band, where QoS is ensured. The user-plane data can be transmitted on either licensed or unlicensed carriers. In this fashion, the crucial information can always be transmitted with QoS guarantee while the unlicensed carriers can provide opportunistic best-effort data transmission enhancements.

Unlike licensed spectrum which is exclusive for the licensed users, unlicensed spectrum is usually shared by multiple unlicensed systems. Therefore, the foremost issue in unlicensed network is to achieve friendly co-existence among multiple unlicensed systems. For

instance, the most successful unlicensed network, Wi-Fi, uses a "listen before talk" (LBT) based MAC protocol, namely, carrier sensing multiple access with collision avoidance (CSMA/CA), to ensure long-term fairness among different users. Thus, to implement LTE-U, it is essential for LTE-U to co-exist in a fair manner with other unlicensed networks especially Wi-Fi. However, LTE and Wi-Fi adopt different MAC protocols that are fundamentally different, i.e., LTE uses scheduling-based MAC for synchronous transmissions while Wi-Fi implements contention-based MAC for asynchronous transmissions. How to allow efficient and friendly co-existence between synchronous LTE-U and asynchronous Wi-Fi and how to analyze the performance of each co-existing network still remain open and beckon for further investigation.

There are several MAC proposals for LTE-U in the literature. In [71] [72], an on/off transmission cycle is introduced where the on period is used for LTE-U transmissions and the off period for Wi-Fi transmissions. The off duration can be randomly selected [71] or dynamically adjusted according to the collected statistics of Wi-Fi activities [72]. For these mechanisms, the LTE-U transmission is not hinged on the instantaneous channel availability and thus may interrupt the ongoing Wi-Fi transmissions. To improve the co-existence performance, the LBT feature is introduced in the MAC design of LTE-U such that the LTE-U node can transmit only if the channel is sensed idle for a certain duration [73] [74]. In [73], an LTE-U node can transmit for a maximum time ratio in one cycle if the channel is sensed idle; or keep silent for the whole cycle, otherwise. An analytical performance study of the duty cycle based protocol is provided in [74]. With a duty cycle based mechanism, LTE-U senses the channel at the specific time in each duty cycle, which makes it difficult for LTE-U to retrieve the channel access due to the elastic feature of Wi-Fi transmissions. Therefore, it is hard to ensure the coexistence performance of the LTE-U system. As such, it is desirable to design a more fair MAC protocol to achieve high performance of LTE-U while ensuring a certain level of Wi-Fi protection.

Our research works [75,76] aim to tackle the above problems by designing a MAC protocol for LTE-U small cells considering both the Wi-Fi protection and LTE channel retrieving.

The protocol design captures the asynchronous feature of Wi-Fi transmissions in a time-slotted MAC frame structure of LTE, and provides tunable parameters that can adjusted according to the throughput requirements of both systems and the desired Wi-Fi protection level. In addition, analytical analysis is also presented to essentially reveal the relationship among different configurable parameters.

2.5 Dissertation Objectives

Based on introduced research topics, challenges and related works, our research logic path on RRM in CA-based LTE-A systems is illustrated in Fig. 2.2.

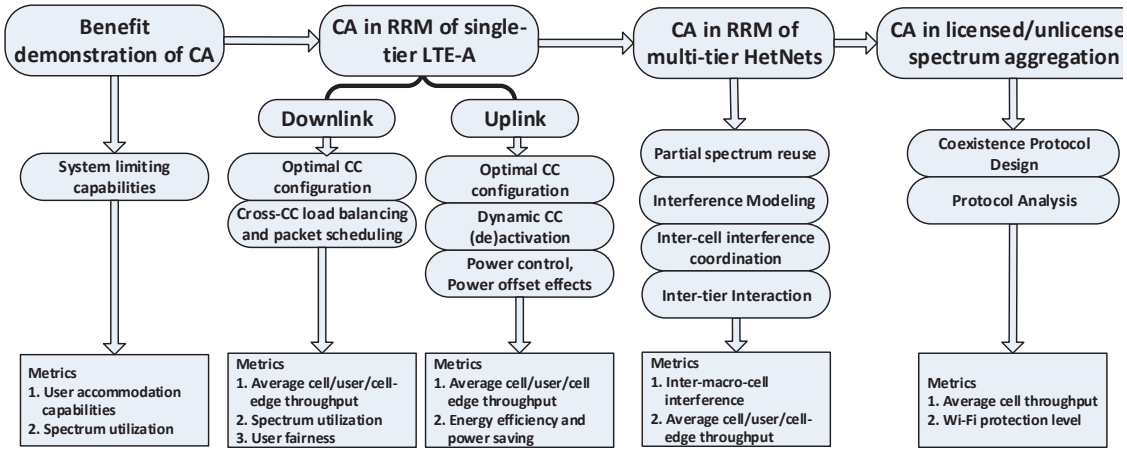


Figure 2.2. Research path of the dissertation on RRM in CA-based LTE-A systems.

The fundamental objective of our research is to fully explore the benefits of CA in the RRM of LTE-A systems. The research starts with the benefit demonstration of CA. Instead of user-centric perspective, we choose the system-centric perspective as the study object. The system user-accommodation capabilities is investigated, i.e., the maximum number of users that can be admitted into the system given the user QoS requirements and traffic descriptors. The LTE users with CA capabilities and the legacy LTE users without the CA capabilities are compared. We would like to know how much performance gain there is in the case when all the users are LTE-A users over that when all the users are LTE ones under different parameter

configuration, and what is the intrinsic reason that results in the performance gain.

After knowing the benefits of CA, we take one step further to study how to take advantage of CA to improve the system performance in single-tier LTE-A system. We focus on the uplink CA since uplink transmissions have more limitations and the downside of CA is non-negligible. We examine the uplink CC selection and power control problems and propose a cross-layer dynamic CC selection and power control strategy with considering the power offset effects.

As the future cellular network will be a heterogeneous networks with different kinds of cells co-existing together, we extend the single-tier CA study to multi-tier CA study, where the interference management between macrocells and small cells and among small cell themselves becomes the theme. As aforementioned, CA enables the partial spectrum usage on the CC resolution. Thus in this step, we study the two-tier HetNets where MCells and FCells share the spectrum in a PSU manner. We model the co-tier and cross-tier interference exploiting the stochastic geometry capturing the locational and temporal randomness of the FCells. As MCells can only influence the FCells indirectly, we model their interaction into Stackelberg game and propose a method to obtain the equilibrium.

Finally, all the above works are considered within the licensed spectrum where the cellular systems have dedicated access to the spectrum resources. CA enables the spectrum aggregation between licensed and unlicensed spectrum. In this way, the system capacity can be significantly increased; meanwhile the system can operate within a unified network structure on both spectrums. Our research focus on how to design the MAC protocol for LTE-U BSs in order to achieve harmony coexistence with other unlicensed systems especially the Wi-Fi systems. The proposed MAC protocol provides tunable coexistence performance with considering the essentially different PHY/MAC specifications between LTE and Wi-Fi.

2.6 Summary

In this chapter, we describe the research topics related to RRM in LTE-A systems with CA. The challenges and related works are also elaborated. Based on the research topics, the research objectives of the dissertation are further outlined. In the next few chapters, we will present our research works on each of the aforementioned topics.

Chapter 3

Equivalent Capacity Analysis in Carrier Aggregation-Based LTE-A Systems

As elaborated in Subsection 2.5, the first step towards knowing the benefits of CA is to compare the network performance of LTE-A with and without the CA features. Most of the existing works on demonstrating the comparison evaluate the network performance based on the user-centric performance, e.g., average user throughput [20, 32]. The theoretical analysis on the system-centric limiting capabilities is still embryonic, which, however, can serve as essential benchmarks for system stability maintenance and Quality of Service (QoS) guarantee.

Thus, in this chapter, we explore one of the primary system limiting capabilities of LTE-A with CA, i.e., the user accommodation capabilities, in the downlink admission control process. The user accommodation capability is analyzed for the legacy LTE and LTE-A users, respectively. The adopted performance metric is *equivalent capacity* (EC) [77], referring to the maximum number of users that can be admitted into the system based on the system bandwidth and user QoS requirements. Specifically, both the LTE and LTE-A users are divided into heterogeneous classes with different QoS requirements (i.e., throughput and loss probability requirements) and traffic descriptors (i.e., active probabilities). Each user class is allocated with a bandwidth weight. Two bandwidth allocation strategies are studied,

namely the fixed-weight strategy and the cognitive-weight strategy, where the bandwidth weights of different classes are pre-fixed in the former while can be dynamically changed in the latter according to the instantaneous load conditions of different classes.

To properly determine the EC in LTE-A systems, the following challenges should be deliberated. First, unlike the wired networks, the channel conditions of the wireless mobile environment are dramatically time-fluctuating due to the complicated propagation environment. Consequently, the assigned bandwidth to satisfy the minimum throughput requirement of a user changes from time to time, making it difficult to determine the EC based on the user throughput requirements. Second, given the system bandwidth and statistics of user traffic, how to determine the EC for each class to maximize the spectrum utilization while satisfying all the user loss probability requirements is a challenging issue. Third, as LTE-A users can transmit with CA, whether the LTE-A users can benefit much from CA over the LTE users needs to be justified. Last but not least, since heterogeneous user classes coexist, the tradeoff among the bandwidth allocation weights for different classes and the criteria therein should be carefully discussed.

Revolving around the above challenges, the contributions of this work are as follows. First, considering the wireless fading statistics, the concept of effective bandwidth [78] is exploited to map the user throughput requirements into bandwidth requirements to provide the users with a probabilistic QoS guarantee. Leveraging the binomial-normal approximation, closed-form expressions of EC are then derived for both LTE and LTE-A users under two bandwidth allocation strategies. Furthermore, a net-profit-maximization problem is formulated to discuss the tradeoff among the bandwidth allocation weights, which combines the factors of operator service profits, user satisfaction and traffic load dynamics. Extensive simulation results are provided to corroborate our analytical ones and demonstrate an interesting discovery that a slightly higher spectrum utilization of LTE-A users than LTE users can result in a significant EC gain when the user traffic is bursty. Moreover, the cognitive-weight strategy outperforms considerably the fixed-weight one in terms of both EC and the achieved net profits.

To the best of our knowledge, this is the first work in literature that gives theoretical analysis on the admission control process in CA-based LTE-A systems. The research outcomes should shed some light not only on theoretically quantifying the benefits of CA but also on the loss-probability-aware bandwidth allocation in admission control process of CA-based LTE-A systems.

3.1 Models and Design Goal

We consider a multi-cell downlink scenario where the base stations (BSs) are deployed following a homogeneous Poisson point process (PPP) [79] with density measure λ_{BS}, i.e., the number of BSs within any given region A with area $|A|$ is a Poisson random variable with parameter $\lambda_{BS}|A|$, and the BSs are uniformly located within A. Denote all the BSs as a set Φ_{BS}. The users are uniformly distributed within region A. Each user will be associated to its nearest BS for service. Under such an association policy, the actual coverage area of a BS becomes a Voronoi cell [80] where any point in a Voronoi cell has a shorter distance to the corresponding BS than to other BSs, as shown in Fig. 3.1. LTE-A and LTE users are con-

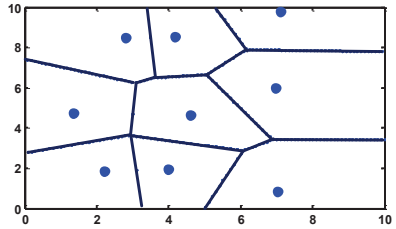


Figure 3.1. Voronoi cells formed by 9 BSs uniformly located within a $10km$x$10km$ area.

sidered for the analysis, respectively. Both LTE and LTE-A users are divided into K classes with different QoS requirements (i.e., the throughput and loss probability requirements) and traffic descriptors (i.e., active probabilities). For each class-k user, the minimum required throughput is r_k^u and the maximum loss probability is δ_k. δ_k refers to the maximum probability that there exist class-k users which are admitted into the system but cannot get any bandwidth when they turn active, i.e., having packets to deliver.

3.1.1 Traffic Generating Model

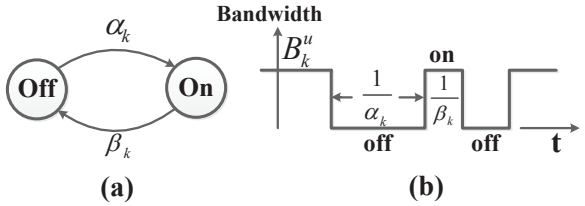


Figure 3.2. On-off traffic generation model of one user.

We consider both voice sources and video sources. Two traffic models which are corroborated by empirical data [81, 82] are exploited, respectively. For the voice sources, we use an on-off traffic source model [81] to describe the dynamics of the user traffic. Each user source is represented by a two-state continuous-time random process with the states "on" and "off", as depicted in Fig. 3.2(a). The state "on" means the user is active and requires a minimum throughput of r_k^u for class-k users, while the user transmits nothing in the state "off". The "on" and "off" intervals can have arbitrarily time distributions but with transition rates denoted as β_k and α_k, respectively. Then the average "on" and "off" period are $1/\beta_k$ and $1/\alpha_k$ as shown in Fig. 3.2(b)). Therefore, the average probability that a class-k user is "on" (i.e., active probability p_k) can be calculated as,

$$p_k = \frac{1/\beta_k}{1/\alpha_k + 1/\beta_k} = \frac{\alpha_k}{\alpha_k + \beta_k},\ k = 1, 2,\ ...,\ K. \tag{3.1}$$

Notice that we consider bursty users satisfying $p_k \ll 1$, e.g., the average *on* period for making voice calls by a user is far smaller than the average *off* period.

Unlike the constant data rate generated by voice sources, one video source has time-varying data rate which cannot be simply modeled as an on-off model. However, according to [82], one class-k video source can be effectively modeled as M_k ($M >> 1$) independent and statistically multiplexed mini-sources with identical constant data rate r_k^{mini}. Each mini-source can be modeled with the above on-off model. Therefore, when the video sources are considered, the equivalent mini-sources can be treated in a similar way with the voice

sources, except that the EC for the mini-sources of the same video user class should be a multiple of M_k. As the main emphasis of this paper lies in the EC comparison between LTE and LTE-A users, we only use voice sources in the following analysis for the simplicity of computation and presentation.

3.1.2 Bandwidth Sharing Model

For bandwidth allocation among different user classes, two strategies are considered, i.e., fixed-weight strategy and cognitive-weight strategy. Under the former strategy, the bandwidth allocation weights ω_{lk} for class-k users are pre-fixed; while under the latter one, the weights can dynamically change with the instantaneous number of active users of each class. Moreover, in the latter strategy, different user classes are endowed with different priorities. Preemptive priority is considered with which the services of some lower-priority users will be interrupted by the higher-priority users when there is no available bandwidth. Thus, the loss probability of one user class is merely affected by the number of users with higher priorities. Without loss of generality, the priorities are set in a descending order with respect to k. In addition, to avoid the spectrum monopoly and evaluate the impact of user dissatisfaction factor on ECs, a maximal bandwidth allocation weight ω_{lk}^{max} is set for each class in the latter strategy. One important application scenario of the latter strategy is the bandwidth allocation for user classes with diverse delay requirements, where the user class with more strict delay requirement will be assigned with higher priority.

For spectrum sharing within one user class, random spectrum access (RSA) method is used. In each transmission time slot, if there are not enough PRBs for every active user in one class, the system randomly chooses a subset of active users to schedule. One scheduled user is randomly assigned with several PRBs according to its QoS requirements. One LTE user can only be assigned with PRBs within one CC; while one LTE-A user can be assigned with PRBs from different CCs. Transmission buffers are not considered here, thus, the packets that are generated in one slot but cannot be transmitted within the same slot are considered

as lost.

3.1.3 Definition of EC

a) Loss-probability-aware ECs. With RSA, the utilization factor ρ_l of the lth CC is defined as,

$$\rho_l := \frac{\sum_{k=1}^{K} M_{lk} p_k B_k^u}{(1-\theta) B_l^{cc}}, \tag{3.2}$$

where M_{lk} is the number of admitted class-k users in lth CC and B_k^u is the effective bandwidth provided by the operator for each class-k user. When ρ_l equals to 1, the spectrum utilization can be maximized, however, the loss probabilities for some class k may exceed the required threshold δ_k. Thus to keep the loss probability of each class under the desired level, ρ_l should be less than 1. To this end, the loss-probability-aware EC for class k in lth CC, denoted as N_{lk}, is defined as the maximum M_{lk} that satisfies the loss probability requirement δ_k, given specified bandwidth allocation strategy and bandwidth weights $\{\omega_{lk}\}$ (or $\{\omega_{lk}^{max}\}$). With the derived EC set $\{N_{lk}\}$, a class-k user will be admitted if the current number of admitted class-k users is less than N_{lk}. As different ω_{lk} (or ω_{lk}^{max}) can result in different ECs, the tradeoff among the bandwidth weights is further discussed in Section 3.4 for both bandwidth allocation strategies.

b) Effective bandwidth from minimum required throughput. As EC is defined on a bandwidth basis, the user's minimum throughput requirement should be mapped into bandwidth requirement to be included in Eq. (3.2). As aforementioned, in a wireless mobile environment, the time-varying channel conditions make it difficult for operator to provide a unified bandwidth to satisfy the minimum throughput requirement r_k^u for all the class-k users at one time. To this end, effective bandwidth is exploited to derive a unified bandwidth B_k^u to provide probabilistic QoS guarantee for all class-k users. In particular, the bandwidth B_k^u should be chosen such that

$$\sup_t \Pr\{B_k^{u,ins}(t) \geq B_k^u\} \leq e \ (0 < e < 1), \tag{3.3}$$

is satisfied, where $B_k^{u,ins}(t)$ denotes the instantaneous minimum required bandwidth to guarantee r_k^u at time t, and e is the upper bound of the QoS violation probability. Eq. (3.3) indicates that given the statistical behaviors of the co-channel interference, the constant bandwidth B_k^u provided by the operator to each class-k user should be no less than $B_k^{u,ins}(t)$ with probability $1-e$. In this way, the effective bandwidth B_k^u provides a bridge between the throughput requirement r_k^u and the EC N_{lk} over the wireless channel statistics. In this work, we mainly focus on the intra-band CA [18], where all the CCs are located in the same frequency band and thus have the same channel fading statistics.

3.1.4 Design Goal

The design goal is to investigate the user accommodation capabilities of CA-based LTE-A systems from the theoretical perspective in order to get deep insight on the benefits of CA. Specifically, the work is first to find the QoS-aware closed-form expressions of EC for both LTE and LTE-A users under different bandwidth allocation strategies. Second, the work is to compare the EC performance between LTE and LTE-A users and between different bandwidth allocation strategies to comprehensively demonstrate in what way and how much the LTE-A users can outperform the LTE users in terms of EC and the economic profits. Third, the research outcomes should shed some lights not only on theoretically quantifying the benefits of CA but also on providing the admissible region for the loss-probability-aware bandwidth allocation in admission control process of CA-based LTE-A systems.

3.2 Equivalent Capacities with Fixed Bandwidth Allocation Weights

The fixed-weight bandwidth allocation strategy is analyzed in this section. Specifically, we first show how to obtain the user effective bandwidth from the throughput requirement, considering the multi-cell co-channel interference. Then, the analysis on a single CC is con-

ducted to find the closed-form $\{N_{lk}\}$-B_l^{cc} relationship, subject to the user equivalent QoS requirements (i.e., B_k^u and δ_k) and traffic descriptors (i.e., active probability p_k). The results are further extended to an LTE-A system with n aggregated CCs, where the ECs are compared between LTE and LTE-A users under the same system setting.

3.2.1 Effective Bandwidth from User Throughput Requirement

For downlink LTE-A systems, as users within the same cell are assigned with orthogonal PRBs, there is no interference among themselves. The co-channel interference is only from other cells that use the same PRBs with the considered cell. According to the LTE-A standard [11], the frequency reuse factor of LTE-A systems is 1, which means each cell operates on the same spectrum and the co-channel interference of the considered cell is the summation of the interference from all the other cells in area A. For a probabilistic analysis, the statistical behaviors of the user SINR (i.e., user SINR distributions) are required. Our previous work [32] has applied the Stochastic geometry [66] to provide tractable probabilistic interference modeling for users in LTE-A systems. Therefore, based on [32], the detailed analysis for user SINR distribution and effective bandwidth is shown as follows.

The channel gain of a user consists of two parts: the path loss and fast fading. The shadowing effect is not considered since the shadowing is shown to be well approximated by the randomness of the Poisson distributed BS locations [83]. This is a strong justification that we can model the locations of BSs into a PPP. Denote the power spectral density (PSD) of the BS transmission power and noise power as P_t and N_0, respectively. Then, the PSD of the received power of a user from BS B is calculated as

$$P_r = P_t H {D_B}^{-\alpha},\ B \in \Phi_{BS}, \tag{3.4}$$

where H is the fast fading channel gain; D_B is the distance between the considered user to

BS B; and α is the path loss exponent. Then the user SINR is calculated as

$$SINR_k = \frac{P_t H {D_{B_0}}^{-\alpha}}{\sum\limits_{B \in \Phi_{BS}^{inf} \backslash B_0} P_t H^{inf} {D_B}^{-\alpha} + n_0}, \tag{3.5}$$

where Φ_{BS}^{inf} denotes the set of interfering BSs in Φ_{BS} that transmit on the same bandwidth with the considered user; $\Phi_{BS}^{inf} \backslash B_0$ means the set of interfering BSs excluding B_0; and H^{inf} denotes the fast fading channel gain between the considered user and the interfering BSs. In this paper, the fast fading between the considered user and the serving BS B_0 is Rayleigh fading, and the fast fading between the considered user and the interfering BSs is generally distributed. Therefore, the probability density function of H is an exponential distribution with parameter μ. For simplicity, μ is set to 1.

To obtain the SINR distribution for one user, we calculate the cumulative probability function (cdf) of $SINR_k$. The probability that $SINR_k$ is larger than a threshold T_k is

$$\begin{aligned} &\mathrm{P}\{SINR_k > T_k\} \\ &= \mathrm{P}\{\tfrac{P_t H {D_{B_0}}^{-\alpha}}{I+n_0} > T_k\} = \mathrm{P}\{H > \tfrac{(I+n_0){D_{B_0}}^{\alpha} T_k}{P_t}\} \\ &where\ I = \sum_{B \in \Phi_{BS}^{inf} \backslash B_0} P_t H^{inf} {D_B}^{-\alpha}. \end{aligned} \tag{3.6}$$

Following the procedure in [32], the final result of $\mathrm{P}(SINR_k > T_k)$ is given directly as

$$\begin{aligned} &\mathrm{P}\{SINR_k \geq T_k\} = 1 - \textstyle\int_0^{+\infty} 2\pi\lambda_{BS} r e^{-\pi\lambda_{BS} r^2} e^{-n_0 r^{\alpha} \frac{T_k}{P_t}} \\ &\qquad \cdot \exp\{-2\pi\theta^{usa}\lambda_{BS}\rho(r, H^{inf}, T_k)\} dr, \\ &where\ \rho(r, H^{inf}, T_k) = -\tfrac{1}{2}r^2 + \tfrac{1}{2}r^2 E_{H^{inf}}\{e^{-T_k H^{inf}} + \\ &(T_k H^{inf})^{2/\alpha}\left[\Gamma(1-\tfrac{2}{\alpha}, 0) - \Gamma(1-\tfrac{2}{\alpha}, T_k H^{inf})\right]\}, \\ &and\ \ \Gamma(s,t) = \textstyle\int_t^{+\infty} x^{s-1} e^{-x} dx. \end{aligned} \tag{3.7}$$

In Eq. (3.7), θ^{usa} is the bandwidth usage probability that one BS transmits on the same bandwidth with the considered user. If the worst case is considered, i.e., every BS transmits

on the same bandwidth with the considered user, the bandwidth usage probability is equal to 1. According to Eq. (3.3), the effective bandwidth B_k^u should be chosen such that

$$\sup_t \mathrm{P}\{B_k^{u,ins}(t) \geq B_k^u\} \leq e \ (0 < e \ll 1), \tag{3.8}$$

that is,

$$\sup_t \mathrm{P}\{r_k^{u,ins}(t) \leq r_k^u\} \leq e \ (0 < e \ll 1), \tag{3.9}$$

where $B_k^{u,ins}(t)$ is the instantaneous required bandwidth of class-k user, and $r_k^{u,ins}(t)$ is the instantaneous achieved throughput when bandwidth B_k^u is assigned. The left hand side of Eq. (3.9) can be rewritten as

$$\begin{aligned} &\sup_t \mathrm{P}\{r_k^{u,ins}(t) \leq r_k^u\} = \mathrm{P}\{B_k^u \log(1 + SINR_k) \leq r_k^u\} \leq e \\ &\Rightarrow \ \ \mathrm{P}\{SINR_k \geq 2^{r_k^u/B_k^u} - 1\} \geq 1 - e \end{aligned} \tag{3.10}$$

Let $T_k = 2^{r_k^u/B_k^u} - 1$. Based on Eq. (3.7), B_k^u can be finalized as the minimum integral multiple of $N_l^{sc}B_l^{sc}$ that satisfies Eq. (3.10), where $N_l^{sc}B_l^{sc}$ is the bandwidth for one PRB, and the constraint of integral multiple is due to the fact that one PRB is the minimum bandwidth allocation unit in LTE-A systems.

3.2.2 Equivalent Capacity for a Single Carrier

For the fixed-weight strategy, ω_{lk} is fixed, thus N_{lk} is only related to the parameters of class k. As aforementioned in Subsection 3.1.3, if the utilization factor ρ_l equals to 1, the loss probability requirements may not be guaranteed. To control the loss probabilities below desired levels $\{\delta_k\}$, the bandwidth assigned to class k should be larger than the average bandwidth requirement of class k, i.e.,

$$B_{lk}^{cc} > N_{lk} p_k B_k^u, \ l = 1, 2, ..., n; \ k = 1, 2, ..., K, \tag{3.11}$$

where $B_{lk}^{cc} = \omega_{lk}(1-\theta)B_l^{cc}$ is the bandwidth assigned to class-k users. ω_{lk} denotes the weight of B_{lk}^{cc}, satisfying $\sum_{k=1}^{K} \omega_{lk} = 1$. To make Eq. (3.11) an equality, we intuitively and heuristically add some multiple of $\sigma_{lk}B_k^u$ to the right hand side of Eq. (3.11), where σ_{lk} is the standard deviation of the number of active class-k users in lth CC [86], thus satisfying $\sigma_{lk} = \sqrt{N_{lk}p_k(1-p_k)}$, i.e.,

$$B_{lk}^{cc} = N_{lk}p_kB_k^u + \Gamma_k\sigma_{lk}B_k^u, \tag{3.12}$$

where Γ_k is a constant which varies with the specified QoS requirement (e.g., loss probability δ_k). Γ_k should increase when the QoS is more strictly defined (i.e., δ_k becomes smaller) while decrease approaching 0 when QoS is made more loose (i.e., δ_k becomes larger). In the following, systematic deductions are given to justify Eq. (3.12) with derived relation between Γ_k and δ_k. Substituting σ_{lk} into Eq. (3.12), we can get

$$B_{lk}^{cc} = (N_{lk}p_k + \Gamma_k\sqrt{N_{lk}p_k(1-p_k)})B_k^u. \tag{3.13}$$

By normalizing the bandwidth B_{lk}^{cc} with B_k^u, we can obtain

$$B_{lk}^{cc,nor} = N_{lk}p_k + \Gamma_k\sqrt{N_{lk}p_k(1-p_k)}, \tag{3.14}$$

where $B_{lk}^{cc,nor} = B_{lk}^{cc}/B_k^u$ is the normalized bandwidth, and $\lfloor B_{lk}^{cc,nor} \rfloor$, i.e., the largest integer smaller than $B_{lk}^{cc,nor}$, indicates the maximum number of class-k users that can be served concurrently by bandwidth B_{lk}^{cc}. With STM method, $B_{lk}^{cc,nor} < N_{lk}$ should hold when $p_k < 1$.

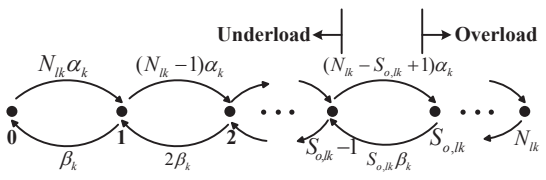


Figure 3.3. Composite traffic model for the system states

In our analysis, all the users are on or off independently. Therefore, the N_{lk} class-k users, each with the traffic generation model described in Subsection 3.1.2, give rise to the composite traffic model in Fig. 3.3. This model is essentially an $(N_{lk}+1)$-state birth-death random process, where state i indicates that there are i active class-k users in the system. Due to the independency among all the users, the transition rate of each user could be added together directly to form the inter-state transition rates as shown in Eq. (3.15). We let $R_{i,j}$ denote the transition rate from state i to state j.

$$\begin{cases} R_{i,i+1} = (N_{lk} - i)\alpha_k, & 0 \le i \le N_{lk} - 1, \\ R_{i,i-1} = i\beta_k, & 1 \le i \le N_{lk}, \\ R_{i,j} = 0, & |i - j| \ne 1. \end{cases} \tag{3.15}$$

The overload state $S_{o,lk}$ is defined for each class k, satisfying $S_{o,lk} = \lceil B_{lk}^{cc,nor} \rceil$. The state $S_{o,lk}$ is the critical state above which losses will occur, i.e, the system is overloaded. We now calculate the steady-state probabilities $\{\pi_{ik}\}$ $(i = 1, ..., N_{lk})$ that the system is in state i. As all the users are independent from each other, the probabilities $\{\pi_{ik}\}$ conform to the binomial distribution, where i out of N_{lk} users are active and the other $(N_{lk} - i)$ are off. Therefore, we have

$$\pi_{ik} = C_{N_{lk}}^{i} {p_k}^{i} (1 - p_k)^{N_{lk} - i}, \ i = 0, 1, \ ..., \ N_{lk}, \tag{3.16}$$

where $C_{N_{lk}}^{i}$ is the number of choices when picking i out of N_{lk}. With $\{\pi_i\}$, the loss probability δ_k can be calculated as the summation of the steady-state probabilities of all the overload states, i.e.,

$$\delta_k = \sum_{i=S_{o,lk}}^{N_{lk}} \pi_{ik}. \tag{3.17}$$

It is well known that when $N_{lk} \gg 1$ and p_k is much smaller than 1, the binomial distribution π_{ik} can be closely approximated by the normal distribution with the mean value $N_{lk}p_k$ and variance $\sigma_{lk}^2 = N_{lk}p_k(1 - p_k)$ [87]. In this work, this condition is satisfied for large number of class-k users with bursty traffic. We convert i and summation to continuous

variable x and integrals, respectively, and then the loss probability δ_k is approximated by

$$\delta_k \approx \int_{S_{o,lk}}^{+\infty} \frac{e^{-(x-N_{lk}p_k)^2/(2{\sigma_{lk}}^2)}}{\sqrt{2\pi}\sigma_{lk}} dx = \frac{1}{\sqrt{\pi}} \int_{\frac{S_{o,lk}-N_{lk}p_k}{\sqrt{2}\sigma_{lk}}}^{+\infty} e^{-y^2} dy, \tag{3.18}$$

where the second step is achieved by replacing x with $\sqrt{2}\sigma_{lk}y + N_{lk}p_k$. Then, integration by parts method is applied to keep the dominant parts of the integration. By multiplying numerator and denominator of the integral by y, we have,

$$\int_z^{+\infty} \frac{ye^{-y^2}}{y} dy = \frac{e^{-z^2}}{2z} - \frac{e^{-z^2}}{4z^3} + \frac{3}{4} \int_z^{+\infty} \frac{e^{-y^2}}{y^4} dy, \tag{3.19}$$

where $z = (S_{o,lk} - N_{lk}p_k)/\sqrt{2}\sigma_{lk}$. When z is greater than 3 (can be easily guaranteed in our setting), the first two items on the RHS of Eq. (3.19) are much larger than the third one. Therefore, we can obtain

$$\int_z^{+\infty} e^{-y^2} dy \approx (\frac{e^{-z^2}}{2z} - \frac{e^{-z^2}}{4z^3}). \tag{3.20}$$

Finally, since $S_{o,lk} \approx B_{lk}^{cc,nor}$, combining (3.18)-(3.20), we derive the expression of loss probability δ_k as,

$$\delta_k \approx \frac{\sigma_{lk} e^{-(B_{lk}^{cc,nor} - N_{lk}p_k)/2{\sigma_{lk}}^2}}{\sqrt{2\pi}(B_{lk}^{cc,nor} - N_{lk}p_k)}. \tag{3.21}$$

Take natural logs for both sides of (3.21), we have

$$\ln(\sqrt{2\pi}\delta_k) = \ln \frac{\sigma_{lk}}{(B_{lk}^{cc,nor} - N_{lk}p_k)} - \frac{(B_{lk}^{cc,nor} - N_{lk}p_k)^2}{2{\sigma_{lk}}^2}. \tag{3.22}$$

The first item on the RHS of (3.22) is neglectable compared with the second, so Eq. (3.22) can be further simplified and rearranged as

$$B_{lk}^{cc,nor} \approx N_{lk}p_k + \sigma_{lk}\sqrt{-2\ln\delta_k - \ln(2\pi)}. \tag{3.23}$$

Now the heuristic thought we proposed in Eq. (3.14) previously is justified with $\Gamma_k = \sqrt{-2\ln\delta_k - \ln(2\pi)}$, and N_{lk} can be solved easily from Eq. (3.23),

$$\begin{aligned} &N_{lk} = \left\lfloor \tfrac{B_{lk}^{cc,nor}}{p_k} - \tfrac{1}{p_k}\left[\sqrt{4\lambda_k(B_{lk}^{cc,nor}+\lambda_k)} - 2\lambda_k\right]\right\rfloor, \\ &where \\ &\lambda_k = {\Gamma_k}^2(1-p_k)/4,\ B_{lk}^{cc,nor} = \omega_{lk}(1-\theta)B_l^{cc}/B_k^u. \end{aligned} \tag{3.24}$$

Therefore, combining Eq. (3.11), (3.14) and (3.24), we can finally conclude that for the lth CC, given the fixed bandwidth weight ω_{lk}, users' effective bandwidth B_k^u, loss probability δ_k and active probability p_k, the normalized bandwidth $B_{lk}^{cc,nor}$ and the EC N_{lk} have the relationship shown in Eq. (3.24) for each class k.

3.2.3 Equivalent Capacity in Multi-Carrier LTE-A Systems

The ECs between LTE users and LTE-A users for each class k are compared in this subsection. n CCs are considered. For the LTE users, as they cannot aggregate PRBs from different CCs, the total EC for class k, denoted as N_k^{LTE}, should be the summation of every N_{lk}. Hence we have,

$$\begin{aligned} &N_k^{LTE} = \sum_{l=1}^{n} N_{lk} \\ &= \sum_{l=1}^{n} \left\lfloor \tfrac{B_{lk}^{cc,nor}}{p_k} - \tfrac{1}{p_k}\left[\sqrt{4\lambda_k(B_{lk}^{cc,nor}+\lambda_k)} - 2\lambda_k\right]\right\rfloor, \\ &where \\ &\lambda_k = {\Gamma_k}^2(1-p_k)/4,\ \Gamma_k = \sqrt{-2\ln\delta_k - \ln(2\pi)}, \\ &B_{lk}^{cc,nor} = \omega_{lk}(1-\theta)B_l^{cc}/B_k^u. \end{aligned} \tag{3.25}$$

For LTE-A users, they can use PRBs from different CCs to transmit concurrently on a wider aggregated virtual bandwidth. The normalized virtual bandwidth for each class k

is $\sum_{l=1}^{n} B_{lk}^{cc,nor}$. Therefore, the total EC for class k, denoted as N_k^{LTE-A}, is given below,

$$N_k^{LTE-A} = \begin{cases} N_k^{LTE}, \ \ if \ \ \left\lfloor \sum_{l=1}^{n} B_{lk}^{cc,nor} \right\rfloor - \sum_{l=1}^{n} \lfloor B_{lk}^{cc,nor} \rfloor < 1; \\ \left\lfloor \frac{\sum_{l=1}^{n} B_{lk}^{cc,nor}}{p_k} - \frac{1}{p_k} \left[\sqrt{4\lambda_k (\sum_{l=1}^{n} B_{lk}^{cc,nor} + \lambda_k)} - 2\lambda_k \right] \right\rfloor, \ otherwise, \end{cases} \tag{3.26}$$

with the same λ_k, Γ_k and B_{lk} defined in Eq. (3.25). From Eq. (3.25) and (3.26), we can observe that the ECs of both LTE and LTE-A users increase when one of the following situations occurs: $i)$ the active probability p_k decreases, $ii)$ the loss probability requirement δ_k is relaxed, or $iii)$ the normalized bandwidth $B^{cc,nor}$ increases. Besides, when $\left\lfloor \sum_{l=1}^{n} B_{lk}^{cc,nor} \right\rfloor - \sum_{l=1}^{n} \lfloor B_{lk}^{cc,nor} \rfloor < 1$, the maximal number of LTE-A users that can be concurrently served is equal to that of LTE users, since even combining the left PRBs of all CCs is not enough to support one more user. In this situation, N_k^{LTE-A} and N_k^{LTE} are equal, which is called the zero-gain situation. However, when it is not the case, more PRBs will be used by LTE-A users for multi-CC transmission, and N_k^{LTE-A} will be larger than N_k^{LTE}, which will be verifed in Section 3.5.

3.3 Equivalent Capacities with Cognitive Bandwidth Allocation Weights

In this section, the cognitive-weight strategy is considered. In order to reduce the analytical complexity, we only conduct analysis on the two-user-class case. Results for cases with more than 2 classes can be deducted following the same method. Similar to Section 3.2, we first derive the closed-from relationship in single-carrier LTE-A systems and then extend the relationship to multi-carrier case where the ECs between LTE and LTE-A users are compared.

3.3.1 Single-Carrier Case

In this section, we consider downlink transmission in a single cell with only two user classes. Without loss of generality, class 1 is assumed with higher priority than class 2. We follow the effective bandwidth mapping in Section 3.2. All the previous notations are applicable in this section. Moreover, we denote ω_{l1}^{max} as the maximal bandwidth weight for class 1 in lth CC, thus $\omega_{l1} \leq \omega_{l1}^{max}$. Then, the maximal number of class-1 users that can be served concurrently by lth CC (denoted as K_{l1}) is

$$K_{l1} = \left\lfloor \frac{B_l^{cc} \omega_{l1}^{max} (1-\theta)}{B_1^u} \right\rfloor. \tag{3.27}$$

Given the number of class-1 users concurrently being served n_{l1}, the maximum number of class-2 users that can be concurrently served by lth CC, denoted as $K_{l2}(n_{l1})$, can be expressed as

$$K_{l2}(n_{l1}) = \begin{cases} \lfloor a(n_{l1}) \rfloor, \ if \ n_{l1} \leq K_{l1}; \\ \lfloor a(K_{l1}) \rfloor, \ if \ n_{l1} > K_{l1}; \end{cases} \tag{3.28}$$

where $a(\cdot)$ is a function defined as

$$a(x) = \frac{B_l^{cc}(1-\theta) - xB_1^u}{B_2^u}. \tag{3.29}$$

As aforementioned, all the users are on or off independently with the traffic generation model described in Subsection 3.1.2. Thus the N_{l1} class-1 and N_{l2} class-2 users can form two independent birth-death random processes with $(N_{l1}+1)$ and $(N_{l2}+1)$ states, respectively, as shown in Fig. 3.3. Having preemptive priority over class-2 users, one class-1 user can get served immediately as long as $\omega_{l1} \leq \omega_{l1}^{max}$ after it is accepted. As a result, the overload state of class 1 is fixed while that of class 2 changes with n_{l1},

$$S_{o,l1} = K_{l1} + 1, \qquad S_{o,l2} = K_{l2}(n_{l1}) + 1. \tag{3.30}$$

Given the loss probability requirement δ_1 and active probability p_1 of class-1 users, the EC N_{l1} can be calculated similarly as Eq. (3.24),

$$\begin{aligned} &N_{l1} = \left\lfloor \tfrac{K_{l1}}{p_1} - \tfrac{1}{p_1}\left[\sqrt{4\lambda_1(K_{l1}+\lambda_1)} - 2\lambda_1\right]\right\rfloor, \\ &where \\ &\lambda_1 = {\Gamma_1}^2(1-p_1)/4,\ \Gamma_1 = \sqrt{-2\ln\delta_1 - \ln(2\pi)}. \end{aligned} \tag{3.31}$$

The calculation of N_{l2} is the most challenging part of the deduction. Different from the fixed-weight strategy, N_{l2} in the cognitive-weight strategy is closely related to the current number of admitted class-1 users (denoted as N_{l1}^{ad}). N_{l1}^{ad} is considered to be $K_{l1} \ll N_{l1}^{ad} \leq N_{l1}$, because the number of admitted users can be much larger than the maximal number of users that can be concurrently served due to the small active probability in the considered scenarios.

Given N_{l1}^{ad}, ω_{l1}^{max} and loss probability requirement δ_2, we still exploit the binomial-normal approximation to get a closed-form relationship between B_l^{cc} and N_{l2}. With the steady-state probabilities $\{\pi_{i1}\}$ and $\{\pi_{i2}\}$, the loss probability of class-2 users can be calculated with two parts,

$$\begin{aligned} \delta_2 = &\underbrace{\sum_{n_{l1}=0}^{K_{l1}} \pi_{n_{l1}1} \sum_{n_{l2}=\lfloor a(n_{l1})\rfloor+1}^{N_{l2}} \pi_{n_{l2}2}}_{A_1} + \underbrace{\sum_{n_{l1}=K_{l1}}^{N_{l1}^{ad}} \pi_{n_{l1}1} \sum_{n_{l2}=\lfloor a(K_{l1})\rfloor+1}^{N_{l2}} \pi_{n_{l2}2}}_{A_2}. \\ &where\ \pi_{n_{l1}1} = C_{N_{l1}^{ad}}^{n_{l1}} {p_1}^{n_{l1}}(1-p_1)^{N_{l1}^{ad}-n_{l1}},\ \pi_{n_{l2}2} = C_{N_{l2}}^{n_{l2}} {p_2}^{n_{l2}}(1-p_2)^{N_{l2}-n_{l2}}, \\ &\quad n_{l1} = 0,1,...,N_{l1}^{ad},\ n_{l2} = 0,1,...,N_{l2}. \end{aligned} \tag{3.32}$$

A_1 is the loss probability of class-2 users when there is no loss from class-1. A_2 is the loss probability when the loss of class 1 occurs, i.e., maximal bandwidth allocation weight ω_{l1}^{max} is reached. Similar as Section 3.2, since $N_{l1}^{ad} \gg 1$, $N_{l2} \gg 1$ and $p_k \ll 1$ $(k = 1, 2)$, the steady-state probabilities $\{\pi_{n_{lk}k}\}$ can be approximated by normal distribution with mean value $N_{l1}^{ad}p_1$ (or $N_{l2}p_2$) and variance $\sigma_{l1}^2 = N_{l1}^{ad}p_1(1-p_1)$ (or $\sigma_{l2}^2 = N_{l2}p_2(1-p_2)$) as shown

below,

$$\begin{aligned}\delta_2 = &\int_{-\infty}^{K_{l1}} \frac{e^{-(x-N_{l1}^{ad}p_1)^2/(2{\sigma_{l1}}^2)}}{\sqrt{2\pi}\sigma_{l1}} \int_{a(x)}^{+\infty} \frac{e^{-(y-N_{l2}p_2)^2/(2{\sigma_{l2}}^2)}}{\sqrt{2\pi}\sigma_{l2}} dydx \quad \Big\} A_1 \\ &+ \int_{K_{l1}}^{+\infty} \frac{e^{-(x-N_{l1}^{ad}p_1)^2/(2{\sigma_{l1}}^2)}}{\sqrt{2\pi}\sigma_{l1}} \int_{a(K_{l1})}^{+\infty} \frac{e^{-(y-N_{l2}p_2)^2/(2{\sigma_{l2}}^2)}}{\sqrt{2\pi}\sigma_{l2}} dydx \quad \Big\} A_2.\end{aligned} \tag{3.33}$$

We first calculate A_1. Similar as Eq. (3.18)-(3.20), we utilize integration by parts to keep the dominant parts of the single integral with respect to variable y. Then the double integral A_3 is turned into single integral with respect to x as

$$A_1 \approx \frac{1}{\sqrt{2}\pi\sigma_{l1}} \left\{ \underbrace{\int_{-\infty}^{K_{l1}} \frac{e^{-\frac{(x-N_{l1}^{ad}p_1)^2}{2{\sigma_{l1}}^2} - \frac{[a(x)-N_{l2}p_2]^2}{2{\sigma_{l2}}^2}}}{\sqrt{2}[a(x)-N_{l2}p_2]/\sigma_{l2}} dx}_{A_3} - \underbrace{\int_{-\infty}^{K_{l1}} \frac{e^{-\frac{(x-N_{l1}^{ad}p_1)^2}{2{\sigma_{l1}}^2} - \frac{[a(x)-N_{l2}p_2]^2}{2{\sigma_{l2}}^2}}}{\sqrt{2}[a(x)-N_{l2}p_2]^3/{\sigma_{l2}}^3} dx}_{A_4} \right\}. \tag{3.34}$$

The approximation holds when $[a(x) - N_{l2}p_2]/(\sqrt{2}\sigma_2) > 3$ for all $x \in (-\infty, K_{l1})$, which can be easily satisfied by our settings. After turning A_1 into single integral, we calculate A_3 as labelled in Eq. (3.34). Re-define $z = a(x) - N_{l2}p_2$ and A_3 can be rewritten as

$$\begin{aligned}A_3 &= \frac{\sigma_{l2}B_2^u}{\sqrt{2}B_1^u} \int_{a(K_{l1})-N_{l2}p_2}^{+\infty} e^{-\frac{[r-N_{l1}p_1-(B_2^u/B_1^u)z]^2}{2{\sigma_{l1}}^2} - \frac{z^2}{2{\sigma_{l2}}^2}}/z \, dz \\ &= \frac{\sigma_{l2}B_2^u}{\sqrt{2}B_1^u} \int_{a(K_{l1})-N_{l2}p_2}^{+\infty} \frac{e^{-C_1z^2-C_2z+C_3}}{z} dz \\ &where \\ &C_1 = \frac{(B_2^u/B_1^u)^2}{2{\sigma_{l1}}^2} + \frac{1}{2{\sigma_{l2}}^2} > 0, \; C_2 = \frac{B_2^u(N_{l1}^{ad}p_1 - r)}{{\sigma_{l1}}^2 B_1^u}, \; C_3 = -\frac{(N_{l1}^{ad}p_1 - r)^2}{2{\sigma_{l1}}^2}, \\ &r = \frac{B_2^u + B_l^{cc}(1-\theta) - N_{l2}p_2B_2^u}{B_1^u} \approx \frac{B_l^{cc}(1-\theta) - N_{l2}p_2B_2^u}{B_1^u}.\end{aligned} \tag{3.35}$$

As C_1 is positive, we can still use integration by parts to only keep the first two items which dominate the integral result,

$$\begin{aligned}&A_3 \approx \frac{\sigma_{l2}B_2^u}{\sqrt{2}B_1^u} \left[\frac{1}{2C_1\eta^2 + C_2\eta} - \frac{4C_1\eta + C_2}{(2C_1\eta + C_2)(2C_1\eta^2 + C_2\eta)^2} \right] e^{-C_1\eta^2 - C_2\eta + C_3}, \\ &where \;\; \eta = a(K_{l1}) - N_{l2}p_2.\end{aligned} \tag{3.36}$$

A_4 can be approximated in the same way and given as

$$A_4 \approx \frac{{\sigma_{l2}}^3 B_2^u}{\sqrt{2}B_1^u}\left[\frac{1}{\eta^3(2C_1\eta+C_2)} - \frac{8C_1\eta^3+3C_2\eta^2}{(2C_1\eta^4+C_2\eta^2)^2(2C_1\eta+C_2)}\right] e^{-C_1\eta^2-C_2\eta+C_3}. \quad (3.37)$$

Now we have derived the first part A_1 of loss probability δ_2 with the summation of A_3 and A_4 as shown in Eq. (3.36) and (3.37). The second part A_2 is easier to calculate, since the two integrals in A_2 are independent with each other and thus can be viewed as the product of two single integrals with respect to x and y, respectively. Applying the integration by parts approximation, we can get the value of A_2 as

$$\begin{aligned} A_2 &\approx \tfrac{1}{\pi}\tfrac{({2z_1}^2-1)}{4{z_1}^3}\tfrac{({2z_2}^2-1)}{4{z_2}^3}e^{-C_1\eta^2-C_2\eta+C_3} \\ where\ \ z_1 &= (K_{l1}-N_{l1}^{ad}p_1)/(\sqrt{2}\sigma_{l1}), \\ z_2 &= [a(K_{l1})-N_{l2}p_2]/(\sqrt{2}\sigma_{l2}). \end{aligned} \quad (3.38)$$

Eq. (3.36)-(3.38) can hold only if both z_1 and z_2 are larger than 3, which can be satisfied by our settings. It can be seen that A_3, A_4 and A_2 all have the same exponential part $e^{-C_1\eta^2-C_2\eta+C_3}$. Let ζ denote the summation of all the coefficients of the exponentials from A_1 and A_2. By taking natural logs for the loss probability δ_2, we can have

$$\ln(\delta_2) = \ln(\zeta) - C_1\eta^2 - C_2\eta + C_3, \zeta > 0. \quad (3.39)$$

The first item of RHS in Eq. (3.39) is neglectable compared with the remaining items, thus Eq. (3.39) can be further simplified as

$$\ln(\delta_2) \approx -C_1\eta^2 - C_2\eta + C_3, \quad (3.40)$$

Finally, by solving N_{l2} from Eq. (3.40), the closed-form relationship between EC N_{l2} and

the system bandwidth B_l^{cc} can be obtained given N_{l1}^{ad}, δ_2, p_2 and ω_{l1}^{max},

$$\begin{aligned} &N_{l2} = \left\lfloor \tfrac{a(K_{l1})}{p_2} - \tfrac{1}{p_2}\left[\sqrt{4\lambda_{2l}(a(K_{l1})+\lambda_{2l})} - 2\lambda_{2l}\right]\right\rfloor, \\ &where \;\; \lambda_{2l} = {\Gamma_{2l}}^2(1-p_2)/4, \\ &\Gamma_{2l} = \sqrt{-2\ln(\delta_2) - \left(K_{l1} - N_{l1}^{ad}p_1\right)^2/{\sigma_{l1}}^2}. \end{aligned} \tag{3.41}$$

From Eq. (3.41), we can observe that the structure of the closed-form relationship of EC N_{l2} is very similar to that of N_{l1} (i.e., Eq. (3.31)), except that $i)$ $a(K_{l1})$ is the maximal number of class-2 users that can be concurrently served by lth CC when ω_{l1} reaches its maximum, and $ii)$ Γ_{2l} is not only related to loss probability δ_2 but also modulated by the parameters of class 1. Furthermore, with the decrease of δ_2 or the increase of ω_{l1}^{max}, λ_{2l} will increase and EC N_{l2} will decrease accordingly, which satisfies the intuitions.

3.3.2 Multi-Carrier Case: LTE vs. LTE-A users

In this subsection, we extend the derived closed-form relationship under the cognitive-weight strategy to the multi-carrier case. A single-cell LTE-A system with n CCs are considered. Similar as Subsection 3.2.3, the ECs become different between LTE users and LTE-A users. For the LTE users, as they cannot use PRBs from different CCs concurrently, the total EC for class k ($k = 1, 2$), denoted as N_k^{LTE}, should be the summation of every N_{lk}, as shown below,

$$\begin{aligned} &N_1^{LTE} = \sum_{l=1}^{n} N_{l1} = \sum_{l=1}^{n}\left\lfloor \tfrac{K_{l1}}{p_1} - \tfrac{1}{p_1}\left[\sqrt{4\lambda_1(K_{l1}+\lambda_1)} - 2\lambda_1\right]\right\rfloor, \\ &N_2^{LTE} = \sum_{l=1}^{n} N_{l2} = \sum_{l=1}^{n}\left\lfloor \tfrac{a(K_{l1})}{p_2} - \tfrac{1}{p_2}\left[\sqrt{4\lambda_{2l}(a(K_{l1})+\lambda_{2l})} - 2\lambda_{2l}\right]\right\rfloor, \\ &where \\ &\lambda_1 = {\Gamma_1}^2(1-p_1)/4, \; \Gamma_1 = \sqrt{-2\ln(\delta_1) - \ln(2\pi)}, \\ &\lambda_{2l} = {\Gamma_{2l}}^2(1-p_2)/4, \;\; \Gamma_{2l} = \sqrt{-2\ln(\delta_2) - \left(K_{l1} - N_{l1}^{ad}p_1\right)^2/{\sigma_{l1}}^2}, \\ &K_{l1} = \left\lfloor \tfrac{B_l^{cc}\omega_{l1}^{max}(1-\theta)}{B_1^u} \right\rfloor. \end{aligned} \tag{3.42}$$

On the other hand, LTE-A users can use PRBs from different CCs to transmit concurrently on a wider aggregated virtual bandwidth. For class 1 with higher priority, the aggregated virtual bandwidth is $\sum_{l=1}^{n} B_l^{cc}(1-\theta)\omega_{l1}^{max}$. Define

$$\begin{aligned} K_1 &:= \left\lfloor \textstyle\sum_{l=1}^{n} B_l^{cc}(1-\theta)\omega_{l1}^{max}/B_1^u \right\rfloor, \\ \hat{a}(x) &:= \left[\textstyle\sum_{l=1}^{n} B_l^{cc}(1-\theta) - xB_1^u\right]/B_2^u. \end{aligned} \tag{3.43}$$

Then the maximal number of class-2 users that can be concurrently served by n CCs when $\sum_{l=1}^{n}\omega_{l1}$ reaches the maximum is $\hat{a}(K_1)$. Thus the total EC of class k ($k=1,2$), denoted as N_k^{LTE-A}, is given as,

$$N_1^{LTE-A} = \begin{cases} N_1^{LTE}, \quad if\ K_1 - \sum_{l=1}^{n} K_{l1} < 1, \\ \left\lfloor \frac{K_1}{p_1} - \frac{1}{p_1}\left[\sqrt{4\lambda_1(K_1+\lambda_1)} - 2\lambda_1\right]\right\rfloor, \ otherwise; \end{cases}$$

$$N_2^{LTE-A} = \begin{cases} N_2^{LTE}, \quad if\ K_1 - \sum_{l=1}^{n} K_{l1} < 1\ and\ \hat{a}(\sum_{l=1}^{n} n_{l1}) - \sum_{l=1}^{n} K_{l2}(n_{l1}) < 1,\ \forall n_{l1} \le K_{l1}, \\ \left\lfloor \frac{\hat{a}(K_1)}{p_2} - \frac{1}{p_2}\left[\sqrt{4\lambda_2(\hat{a}(K_1)+\lambda_2)} - 2\lambda_2\right]\right\rfloor, \ otherwise; \end{cases}$$

$$\begin{aligned} &where \\ &\lambda_1 = {\Gamma_1}^2(1-p_1)/4,\ \Gamma_1 = \sqrt{-2\ln(\delta_1) - \ln(2\pi)}, \\ &\lambda_2 = {\Gamma_2}^2(1-p_2)/4,\ \ \Gamma_2 = \sqrt{-2\ln(\delta_2) - \left(K_1 - \textstyle\sum_{l=1}^{n} N_{l1}^{ad}p_1\right)^2/{\sigma_1}^2}, \\ &{\sigma_1}^2 = \textstyle\sum_{l=1}^{n} N_{l1}^{ad}p_1(1-p_1). \end{aligned} \tag{3.44}$$

Under this strategy, the zero-gain situation will hardly occur since its occurring conditions are much harder due to that the bandwidth weight of class-2 users can dynamically change in a cognitive manner as the secondary users do in the cognitive radio networks.

3.4 Net-Profit Maximization Under Different Bandwidth Allocation Strategies

In previous sections, we have derived the closed-form expressions of ECs considering user loss probability requirements for both LTE and LTE-A users under the fixed-weight and cognitive-weight strategies. In this section, we discuss the economic tradeoff among the bandwidth allocation weights of different user classes via a net-profit-maximization problem.

The decision of the bandwidth allocation weights should concern a combination of factors. Among these factors, operator profits, user satisfaction and the dynamic traffic conditions of different user classes are the three most important ones. On one hand, the operators tend to allocate more bandwidth to the user class that can bring higher profits per PRB, i.e., short-term profits. On the other hand, only maximizing the short-term profits may incur undesired non-neglectable user dissatisfaction for some kinds of user classes, which may in turn hurt the operators' long-term benefits. Therefore, the optimal decision of the bandwidth allocation weights should balance the operators' short-term profits and the satisfaction of all the user classes. Moreover, to enhance the bandwidth utilization, the weight decision should be conducted dynamically to adapt to the time-varying traffic conditions of different user classes.

To formulate the tradeoff into an optimization problem, we consider that the average number of users for each user class k in a single cell is constant within a certain time period τ but changes from period to period. For a particular time period τ, denote the average number of class-k users in the cell as $N_k(\tau)$. The $N_k(\tau)$ users can be on or off following the traffic generation model in Subsection 3.1.2. The profit per PRB that the operator will get from class-k users is denoted as $G_{k,PRB}$. Our objective is to maximize the net benefits that the operator can get with considering the user satisfaction factor for each period τ.

3.4.1 Fixed-Weight Bandwidth Allocation Strategy

For the fixed-weight strategy, the optimization problem for LTE users is given as

$$\begin{aligned} \max_{\omega_{lk}} \ & \sum_{k=1}^{K} \left[\min\{N_k(\tau), N_k^{LTE}\} p_k \frac{B_k^u}{B_l^{sc} N_l^{sc}} G_{k,PRB} - \chi_k \max\{N_k(\tau) - N_k^{LTE}, 0\} \right] \\ s.t. \quad & equation\ (3.25); \\ & l = 1, 2, ...n; \quad k = 1, 2, ..., K, \end{aligned} \tag{3.45}$$

where $\min\{N_k(\tau), N_k^{LTE}\}$ is the actual number of admitted class-k users and $\max\{N_k(\tau) - N_k^{LTE}, 0\}$ is the average number of class-k users that are rejected to get into the system. Thus $\min\{N_k(\tau), N_k^{LTE}\} p_k \frac{B_k^u}{B_l^{sc} N_l^{sc}}$ is the average number of PRBs that are occupied by class-k users within period τ. χ_k is a weighting parameter to adjust the relative importance of class-k users' satisfaction over short-term profits. The optimization can be simplified by introducing auxiliary variables $\phi_k = \min\{N_k(\tau), N_k^{LTE}\}$ and $\varphi_k = \max\{N_k(\tau) - N_k^{LTE}, 0\}$,

$$\begin{aligned} \max_{\omega_{lk}} \ & \sum_{k=1}^{K} \left[\phi_k p_k \frac{B_k^u}{B_l^{sc} N_l^{sc}} G_{k,PRB} - \chi_k \varphi_k \right] \\ s.t. \quad & equation\ (3.25), \ \ \forall k = 1, 2, ..., K; \\ & \phi_k \le N_k(\tau), \quad \forall k = 1, 2, ..., K; \\ & \phi_k \le N_k^{LTE}, \quad \forall k = 1, 2, ..., K; \\ & \varphi_k \ge N_k(\tau) - N_k^{LTE}, \quad \forall k = 1, 2, ..., K; \\ & \varphi_k \ge 0, \quad \forall k = 1, 2, ..., K; \\ & 0 \le \omega_{lk} \le 1. \end{aligned} \tag{3.46}$$

The optimization problem (3.46) can be transformed into a quadratic constrained linear optimization problem, which can be effectively solved by the interior-point algorithm [88]. Similarly, for LTE-A users, the optimization problem is the same as (3.46) with replacing Eq. (3.25) with Eq. (3.26).

3.4.2 Cognitive-Weight Bandwidth Allocation Strategy

For the cognitive-weight strategy, the optimization problem for LTE users is formulated similarly as

$$\begin{aligned}
&\max_{\omega_{l1}^{max}} \sum_{k=1}^{2} \left[\min\{N_k(\tau), N_k^{LTE}\} p_k \tfrac{B_k^u}{B_l^{sc} N_l^{sc}} G_{k,PRB} - \chi_k \max\{N_k(\tau) - N_k^{LTE}, 0\}\right] \\
&s.t. \quad \textit{equation (3.42)}; \\
&\quad N_{l1}^{ad} = \min\left\{N_1(\tau), N_1^{LTE}\right\} \omega_{l1}^{max} / \sum_{l=1}^{2} \omega_{l1}^{max}, \ \forall l = 1, 2, ..., n.
\end{aligned} \tag{3.47}$$

The main difference of optimization problem (3.47) from (3.45) is the decision of N_2^{LTE}. In (3.47), N_2^{LTE} is closely related to the number of admitted class-1 users in each CC (i.e., N_{l1}^{ad}) and N_{l1}^{ad} is further limited by N_1^{LTE}; while in (3.45), N_2^{LTE} is only related to its own weight ω_{l2}, which is fixed for a certain τ. Therefore, although the optimal weights are determined for both strategies, respectively, the cognitive bandwidth allocation has stronger adaptability in capturing the time-varying traffic demands of different users, thus having higher bandwidth utilization. Here, we consider that all the admitted LTE users in the same class are assigned to different CCs in proportion to ω_{l1}^{max}. The optimization can be simplified by similarly introducing auxiliary variables $\phi_k = \min\{N_k(\tau), N_k^{LTE}\}$ and $\varphi_k = \max\{N_k(\tau) - N_k^{LTE}, 0\}$ as

$$\begin{aligned}
&\max_{\omega_{l1}^{max}} \sum_{k=1}^{2} \left[\phi_k p_k \tfrac{B_k^u}{B_l^{sc} N_l^{sc}} G_{k,PRB} - \chi_k \varphi_k\right] \\
&s.t. \quad \textit{equation (3.42)}, \ \forall k = 1, 2; \\
&\quad \phi_k \le N_k(\tau), \quad \forall k = 1, 2; \\
&\quad \phi_k \le N_k^{LTE}, \quad \forall k = 1, 2; \\
&\quad \varphi_k \ge N_k(\tau) - N_k^{LTE}, \quad \forall k = 1, 2; \\
&\quad \varphi_k \ge 0, \quad \forall k = 1, 2; \\
&\quad N_{l1}^{ad} = \phi_1 \, \omega_{l1}^{max} / \sum_{l=1}^{2} \omega_{l1}^{max}, \ \forall l = 1, 2, ..., n; \\
&\quad 0 \le \omega_{l1}^{max} \le 1.
\end{aligned} \tag{3.48}$$

The optimization problem (3.48) is still solvable following [88]. For LTE-A users, the optimization problem is the same as (3.48) except that i) Eq. (3.42) is replaced by Eq. (3.44); ii) the equation about N_{l1}^{ad} is replaced by $\sum_{l=1}^{n} N_{l1}^{ad} = \phi_1$ since all the CCs can be aggregated as a virtual band and it is not necessary to know the specific value of each N_{l1}^{ad}.

3.5 Performance Evaluation

To validate our proposed closed-form expression of EC, system-level simulations are conducted with Matlab for downlink transmissions in a single cell. Furthermore, the ECs are compared between LTE users and LTE-A users for both fixed-weight and cognitive-weight strategies. Finally, the results of the formulated optimization problems are presented to illustrate the economic advantages of CA and the cognitive bandwidth allocation strategy. The main simulation parameters are listed in Table 3.1 and Table 3.2, where the effective bandwidth is measured by the number of PRBs per user, i.e., P_k^u.

Table 3.1 Simulation Parameters I

Parameter Group 1	**Values**
transmission power spectrum density, P_0	0.2 W/Hz
noise power spectrum density, N_0	10^{-9} W/Hz
path loss exponent, a	3
cell radius, R_{cell}	$500m$
standard variance of shadowing X	$8dB$
Rayleigh scale parameter, σ	$0.5dB$
Subcarrier bandwidth, B_l^{sc}	15kHz
Number of subcarriers per PRB, N_l^{sc}	12

3.5.1 Effective Bandwidth

We first show how the effective bandwidth B_k^u changes with the minimum throughput requirement r_k^u in Fig. 3.4 via *Monte Carlo* simulations. The Parameter values in Table 3.1 are used. It can be seen that for all the simulated values of e, B_k^u presents a non-decreasing trend with increasing r_k^u. This is because the system has to spend more bandwidth to guarantee

Table 3.2 Simulation Parameters II

Parameter Group 2	Class-1 Values	Class-2 Values
Number of CCs, n	1∼5	1∼5
CC bandwidth, B_l^{cc}	20MHz	20MHz
Number of PRBs per user, P_k^u	8∼20	1∼10
Loss probability, δ_k	$10^{-5} \sim 10^{-1}$	$10^{-6} \sim 10^{-3}$
Active probability, p_k	0.02∼0.2	0.02∼0.2
Total percentage of GBs, θ	0.1	0.1
Normalized profit per PRB, $G_{k,PRB}$	3	2
Satisfaction weighting factor, χ_k	0.075	0.05

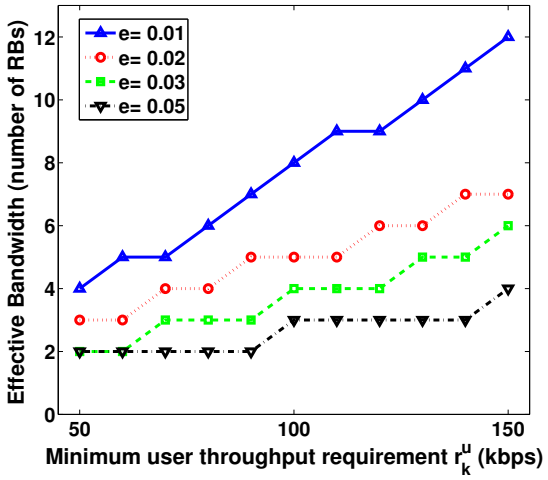


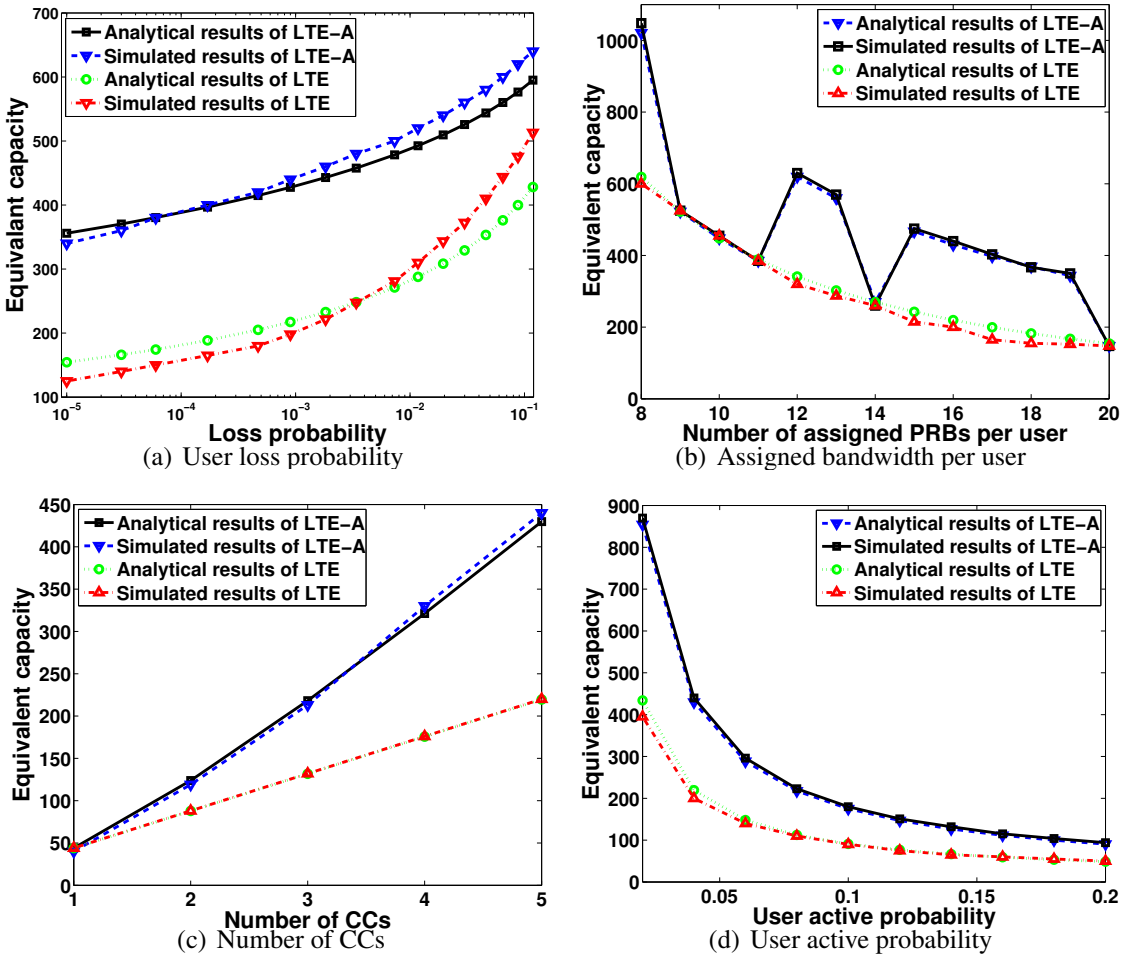
Figure 3.4. Effective bandwidth vs. r_k^u under different e values.

a larger minimum throughput requirement given e and the wireless channel statistics. For each curve, the flat part occurs, because a PRB is the minimum bandwidth allocation unit. Besides, we can observe that the effective bandwidth increases faster with smaller e. This is explained as follows. According to Eq. (3.10), e is positively correlated with r_k^u/B_k^u. Therefore, if r_k^u increases by a fixed ratio, B_k^u has to increase by larger ratio for smaller e, that is, for smaller e, the slope of the r_k^u-B_k^u curve is larger.

3.5.2 Fixed-Weight Strategy

In this subsection, we evaluate the EC performance of the fixed-weight strategy. We first consider single user class to show how the ECs N^{LTE} and N^{LTE-A} vary with different parameters. The class-1 values of Parameter Group 2 in Table 3.2 are used. Both analytical

and simulated results are shown in Fig. 3.5, where the four evaluated parameters are loss probability δ_1, the number of assigned PRBs P_k^u, the number of aggregated CCs n, and active probability p_1. The default values are p_1=0.04, n=5, P_1^u=16, $\delta_1 = 10^{-3}$. We have the following observations.



(a) User loss probability
(b) Assigned bandwidth per user
(c) Number of CCs
(d) User active probability

Figure 3.5. Fixed-weight strategy: ECs under 4 different parameters

i) It can be seen that the analytical and simulated results match well under all parameters for both LTE-A and LTE users, except when the loss probability is higher than 0.01 in Fig. 3.5(a). The reason is that when the loss probability is very high (e.g., larger than 0.01), the third item of Eq. (3.19) becomes non-negligible since z is very close to or even lower than

3, making the approximation inaccurate.

ii) For all subfigures, the ECs of LTE-A users surpass those of LTE users significantly except in the zero-gain situations. The gain (i.e., the ratio of ECs between LTE-A and LTE users) comes from the semi-usage PRBs mentioned in Subsection . Given that these PRBs only account for a very small portion of the whole transmission bandwidth, the notable gain can be explained as follows. Recall Eq. (3.25) and (3.26), due to the introduction of δ_k, EC increases nonlinearly and faster when the normalized bandwidth increases. Take the default settings of Fig. 3.5 as an intuitive example: The 5 CCs can concurrently serve at most 30 LTE users or 31 LTE-A users. When there are 31 LTE users in the system, the loss probability is p^{31}. When there are 32 LTE-A users, the loss probability is p^{32}. As $p \ll 1$, p^{32} is much smaller than p^{31}. Hence, many more LTE-A users can be admitted into the system to make the two loss probabilities the same, which causes the gain to be around 2.

iii) The gain increases as the number of PRBs per user or the number of aggregated CCs increases (except for the zero-gain situation), but stays unchanged with the active probability. This is because the portion of the semi-usage PRBs becomes larger (except the zero-gain situation) with the increase of the number of PRBs per user or CCs, resulting in a larger spectrum utilization with LTE-A users, while the utilization stays the same for different active probabilities.

iv) The zero-gain situations occur in Fig. 3.5(b) when the number of assigned PRBs per user is equal to 9-11,14 or 20. The reason is as follows. Based on the parameter settings, there are totally 100 PRBs per CC. If we take $P_1^u = 9$ as an example, at most 11 users can be concurrently served in each CC. When $n = 5$, the left 5 PRBs (one from each CC) are still not enough to support one more user cooperatively. As a result, the bandwidth utilization remains the same for LTE and LTE-A users, leading to zero gain.

Besides, a two-user-class case is further simulated to exhibit the relationship between the ECs of two user classes under different bandwidth weights, as shown in Fig. 3.6, where LTE and LTE-A users are considered, respectively. The default parameter values of class 1 are the same as the above single-class case; and those of class 2, i.e., P_2^u, p_2 and δ_2 are set to be 15,

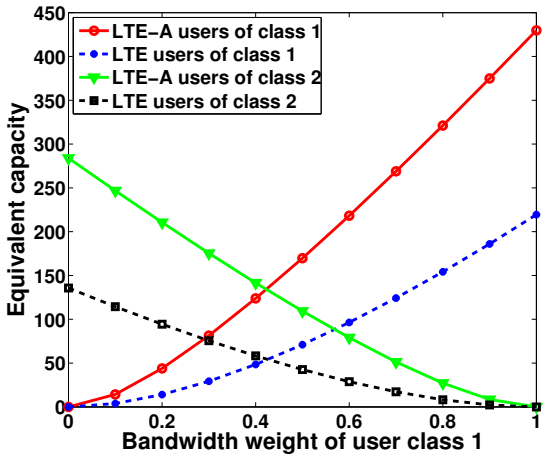


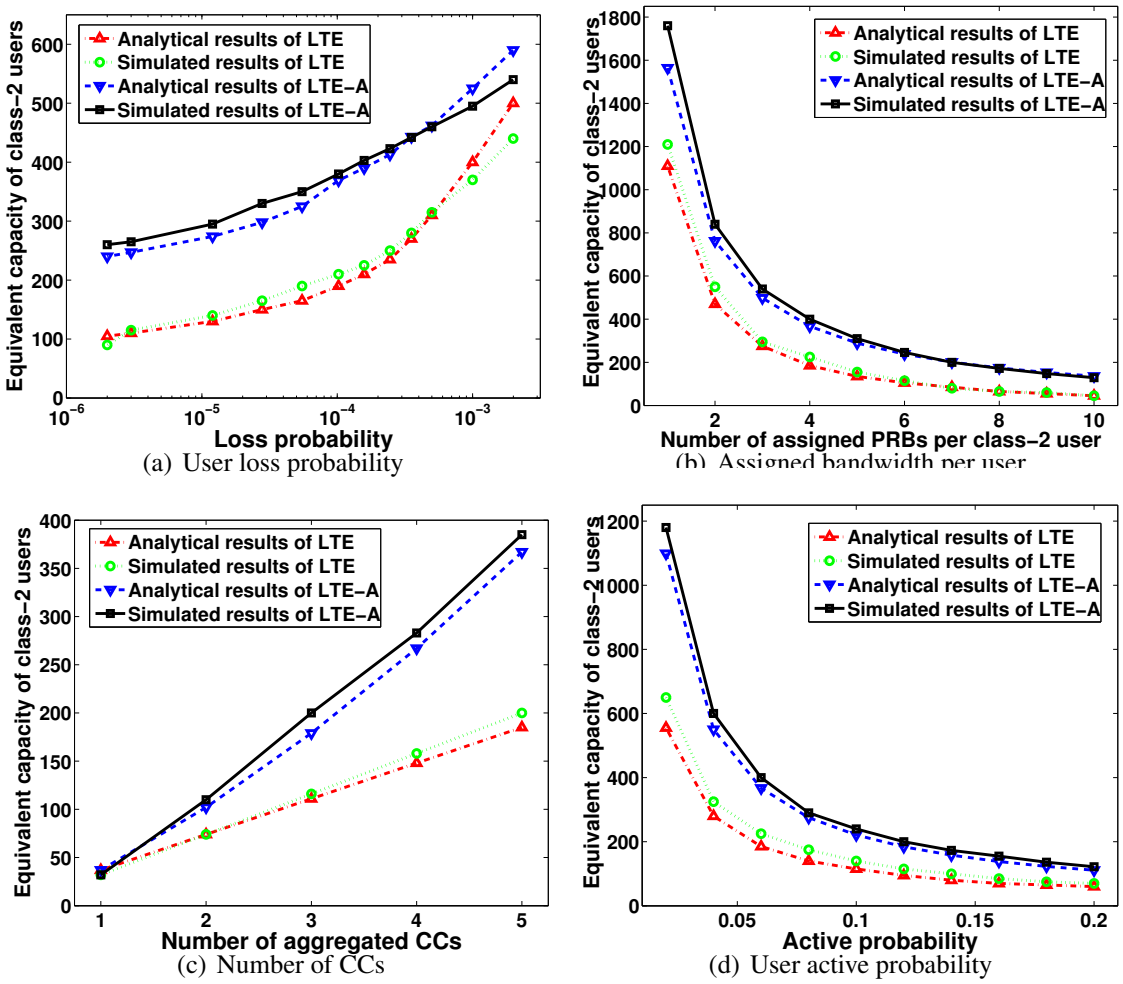
Figure 3.6. Fixed-weight strategy: The EC relation between 2 user classes

0.06 and 10^{-4}, respectively. From Fig. 3.6, we can observe that the ECs of the two classes present a strong negative correlation and either achieves the maximum when assigned with the whole transmission bandwidth.

3.5.3 Cognitive-Weight Strategy

In this subsection, we evaluate the EC performance for the cognitive-weight strategy with two user classes. As the main difference from the previous strategy is the EC of the lower-priority user class (i.e., class 2), we only present the ECs of class-2 users with different changing parameters. As shown in Fig. 3.7-3.8, both analytical and simulated results are given and the Parameter Group 2 in Table 3.2 is used. The default settings are: 1) $\delta_1 = 10^{-4}$, $\delta_2 = 10^{-3}$, $P_1^u = 7$, $P_2^u = 5$, $n = 4$, $p_1 = 0.02$, and $p_2 = 0.06$; 2) $\omega_{l1}^{max} = 0.8$ and the actually admitted number of class-1 users N_{l1}^{ad} is set to be 90% of the EC N_{l1}.

First, it can be observed that the analytical and simulated results match well in both Fig. 3.7 and 3.8 except two cases: when the loss probability is larger than 10^{-3} (Fig. 3.7(a)) and when ω_{l1}^{max} is less than 0.3 (Fig. 3.8). These relatively large gaps are expected and can be explained with the similar reason as in the fixed-weight strategy: when $\delta_2 > 10^{-3}$ or $\omega_{l1}^{max} < 0.3$, the condition that z_2 in Eq. (3.38) should be larger than 3 cannot hold, and thus our approximation method becomes inaccurate.



(a) User loss probability

(b) Assigned bandwidth per user

(c) Number of CCs

(d) User active probability

Figure 3.7. Cognitive-weight strategy: ECs of class-2 users under different parameters.

Besides, the EC gain of the class-2 LTE-A users over the LTE users increases when the number of CCs, the number of assigned PRBs per user or ω_{l1}^{max} increases, while stays almost unchanged with the active probability. This can be explained with the same reason as in the fixed-weight strategy in terms of bandwidth utilization. Moreover, no zero-gain situations occur in Fig. 3.7(b) when P_2^u changes. This is because for cognitive-weight strategy, the conditions for the zero-gain situation are much harder to achieve since the parameters of both user classes are involved therein, as indicated in Eq. (3.44).

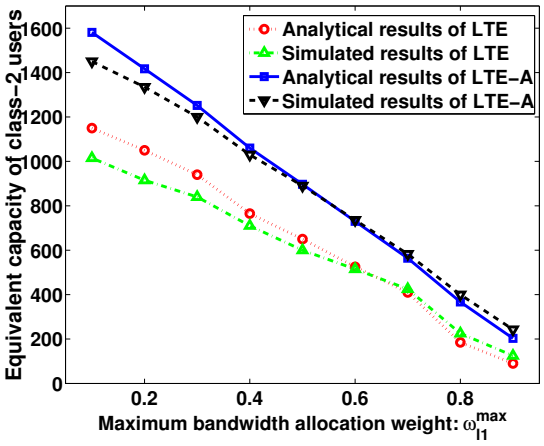


Figure 3.8. Cognitive-weight strategy: ECs of class-2 users with changing ω_{l1}^{max}.

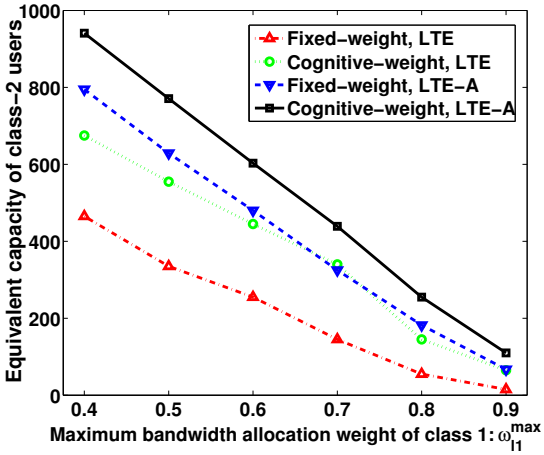


Figure 3.9. EC comparison between two strategies.

3.5.4 Performance Comparison between Two Strategies

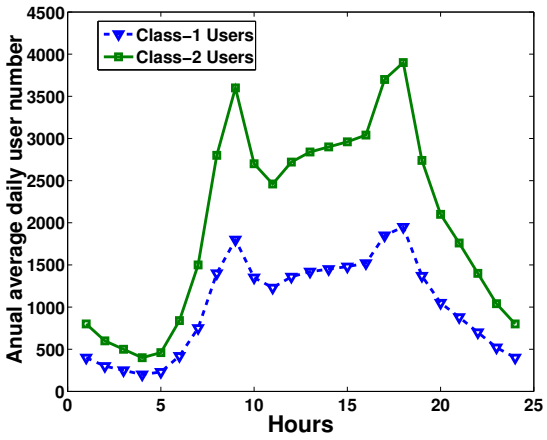


Figure 3.10. Annual average hourly number of users per cell in the tested cell.

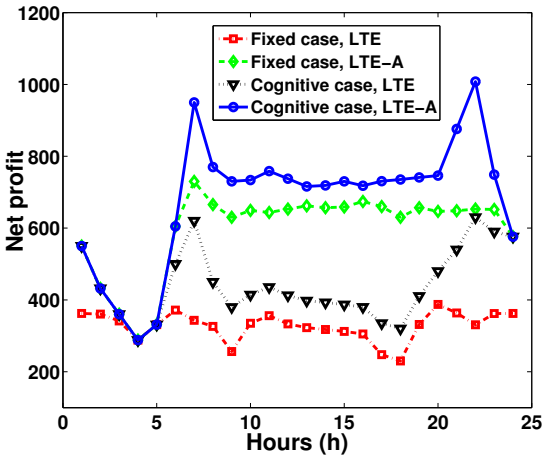


Figure 3.11. Net profits comparison between two strategies.

To compare the performance between the fixed-weight and cognitive-weight strategies, Fig. 3.9 and Fig. 3.10-3.11 are shown respectively in the respect of EC and the achieved normalized operator profits. The Parameter Group 2 in Table 3.2 is used with the same default settings as Subsection 3.5.3 except $n = 5$.

Fig. 3.9 shows how class-2 ECs under different strategies change with the allocation weight of class-1 users. N_{l1}^{ad} is set to the full EC N_{l1}. It can be found that the ECs under the cognitive-weight strategy are considerably higher than those under the fixed-weight strategy for both LTE and LTE-A users. For some weight values, the ECs of LTE users under the cognitive-weight strategy are even close to those of LTE-A users under the fixed-weight one. This is because the bandwidth weight of class-2 users under the cognitive-weight strategy can dynamically change in a cognitive manner as the secondary users do in the typical cognitive radio networks. As a result, more class-2 users can be concurrently served on average, and thus larger EC will be obtained under the same loss probability requirement.

Furthermore, we compare the optimal net profits obtained from the utility-maximization problems formulated in Section 3.4. The scales of the collected annual average hourly traffic in [89] are used to generate the hourly average number of users per cell in Fig. 3.10. The hourly optimal net profits are presented in Fig. 3.11. We can observe that the cognitive-weight strategy outperforms the fixed-weight one significantly for most of the time. The gain comes from that the class-2 ECs under the cognitive-weight strategy are larger when both strategies have the same the class-1 ECs. As shown in Fig. 3.11, for the LTE-A users, when the traffic load is light (e.g., 1-6a.m.), all the users in the cell can be admitted into the system, thus the net benefits under two strategies are the same. As the traffic load increases (e.g., 6-7a.m.), the cell under the fixed-weight strategy will be first saturated and the net profits will be affected by the increasing user dissatisfaction from the rejected users. Since the cell in the cognitive-weight strategy has larger EC, the corresponding net profits will be higher. However, when there are too many users in the cell (e.g., 7-9a.m.), user dissatisfaction will have larger impact on the net profits, thus leading to a profit reduction. The curves of the LTE users can be explained similarly.

3.6 Summary

In this work, we have studied the EC performance of LTE-A systems with CA for LTE and LTE-A users under two bandwidth allocation strategies. The concept of effective bandwidth has been introduced to map the user throughput requirement into the bandwidth requirement with considering the wireless channel statistics. Then, closed-form expressions of EC have been derived with the binomial-normal approximation for both kinds of users under both strategies. We have further formulated a net-profit-maximization problem to investigate the tradeoff among the bandwidth weights for heterogeneous user classes. Finally, extensive simulations have demonstrated the accuracy of our analytical results and shown: $i)$ with only a small increase in the spectrum utilization, LTE-A users can have considerably higher EC than LTE users when the user traffic is bursty; and that $ii)$ the cognitive-weight strategy performs better than the fixed-weight one due to stronger adaptability to the traffic load conditions. For the future work, we will investigate the EC performance of LTE-A systems in a multi-cell scenario.

Chapter 4

Cross-Layer Carrier Selection and Power Control for LTE-A Uplink with CA

There have been abundant research works related to CA-based LTE-A systems. Many studies focus on downlink RRM in LTE-A systems. However, these works can not be directly applied to the uplink CA due to some significant differences. First, the user power constraint is usually a main limitation on the RRM performance. When a user reaches its maximum transmission power, it may not be possible to improve the throughput even if more component carriers (CCs) are allocated to the user. Second, multi-CC transmission in CA can increase the Peak-to-Average Power Ratio (PAPR) and inter-modulation power consumption [49], which further lead to a non-neglectable reduction in user's maximum transmission power. These effects, referred to as power offset effects in this chapter, degrade the user performance inevitably. Thus, it is essential and challenging to address the RRM issues in uplink CA.

There are some recent works [39, 40, 50] dealing with the above issues in uplink CA. However, these works have not considered how the time-variabilities of either CC load conditions or the offset effects impact the RRM performance in uplink CA. In fact, the RRM decisions should be tightly related to the current CC load conditions in order to efficiently utilize the limited user transmission power; and the user power offset should vary with the

number of CCs and the instantaneous resource block occupation conditions. Therefore, it is critical to involve the time-varying features into the RRM strategy design.

In this chapter, we incorporate the time-varying CC load conditions and the infamous power offset effects into the RRM framework, and explore the uplink CC selection and power control process in a single cell. A cross-layer joint CC selection and power control strategy is proposed, which can significantly improve the average user throughput by maximizing the power utilization of the newly admitted user. Specifically, the contributions of this work are fourfold. First, given the current CC load conditions, we put forward a novel quantitative estimation method to predict the average number of RBs that one newly admitted user can get from each CC, leveraging the fairness property of the Layer-2 packet scheduling. Second, the power offset is modeled as a function of the number of assigned CCs and occupied RBs in each CC for each user. Third, based on the estimation results and power offset model, an optimization problem is formulated and solved to find the optimal CC selection decision and power allocation values, such that the user power utilization can be maximized. Finally, comprehensive simulations are conducted to validate the effectiveness of the proposed estimation method and show the performance gain over the existing strategy.

4.1 System Model

We consider the uplink of a single-cell LTE-A system with n aggregated CCs. Users arrive in the system following a Poisson process with parameter λ and are uniformly distributed across the cell. Each user i has a payload of P_i bits to transmit. Denote the maximum transmission power in dBm and the subset of CCs assigned to user i as $P_{i,max}$ and $\mathbb{C}_i$ ($i \in 1, 2, ...$), respectively. As user equipment (UE) power limitation due to multi-CC transmission is the main emphasis in this work, only LTE-A users are involved in the following discussions.

For Layer-3 CC selection process, to achieve better performance in terms of throughput and user fairness, load balancing is considered in this work to guarantee all the CCs to be equally loaded. A simple yet effective load balancing strategy is adopted, where the CC with

the least number of users will be prioritized to bear the newly admitted user. As a result, the difference between the numbers of users in different CCs can be kept under a relatively low level at any time. The strategy is referred to as Least-First in this chapter.

The performance of Layer-2 packet scheduling (PS) is tightly coupled with the spectrum access technologies. Single-Carrier Frequency Division Multiple Access (SC-FDMA) is selected here for LTE-A uplink with CA, which is beneficial to users in the point of power consumption due to its low PAPR compared to its alternative, i.e., OFDMA. Under SC-FDMA technology, the allocated RBs to a UE must be contiguous. A commonly used scheduler, namely Proportional Fair (PF), is adopted independently in each CC to maintain fairness among UEs within a single CC [90]. The basic rationale of PF scheduling is to first select the UE that maximizes a given metric and expand its bandwidth until another UE has a higher metric value on the vicinal RB(s). With PF, UEs in one single CC can have approximately equal long-time average throughput.

Layer-1 power control is implemented according to the formula standardized in [91]. The optional closed-loop regulations are not considered here. If we denote the number of RBs occupied by user i in jth CC as $N_{i,j}^{RB}$, the transmission power that user i spends on CC j (denoted as $P_{i,j}$) can be expressed as follows,

$$P_{i,j} = \min\{P_{i,j}^{max},\ P_0 + 10\log_{10}(N_{i,j}^{RB}) + \alpha L_i\}\ (dBm), \tag{4.1}$$

where P_0 and α are CC-specific power control parameters. In this study, we focus on intra-band CA where all the aggregated carriers lie in the same frequency band and have same P_0 and α. L_i is the path loss in dB due to slow fading,

$$L_i = 10\beta\log_{10}(D_i) + X\ (dB), \tag{4.2}$$

where β is the path loss exponent, D_i is the distance between UE i and its associated base station, and X is a normal-distributed random variable. Note that equation (4.1) is only a raw power allocation plan on each CC. When multiple CCs are assigned to user i, the total

planned power may exceed the maximum available power, thus leading to the necessity to further scale all the power values $P_{i,j}$ ($j \in \mathbb{C}_i$).

The estimated power offset when user i is transmitting on multiple CCs concurrently is denoted as $P_{i,offset}$ in dB. $P_{i,offset}$ depends on many factors such as the number of assigned CCs and allocated RBs in each CC, modulation and coding schemes and so forth, making it a complicated issue [49]. In this work, we model $P_{i,offset}$ as a function of the number of assigned CCs (i.e., $|\mathbb{C}_i|$) and RBs in each CC as below, instead of a constant in [50],

$$P_{i,offset} = (|\mathbb{C}_i| - 1 + \theta \sum_{j \in \mathbb{C}_i} N_{i,j}^{RB}) P_{back}\ (dB), \tag{4.3}$$

where θ ($\ll 1$) and P_{back} are CC-specific constants. Note that $P_{i,offset}$ is not a power value but a scaling ratio in dB.

Shannon formula is used for physical layer rate estimation[1]. For each user i, its throughput on RB c (denoted as $R_{i,c}$) is achieved by,

$$R_{i,c} = W_{RB} \log_2 \left\{ 1 + \frac{10^{[(P_{i,c} - L_i) - 30]/10}}{N_0} \right\}\ (bit/s), \tag{4.4}$$

where W_{RB} is the bandwidth per RB, $P_{i,c}$ is the power in dBm of user i on RB c, and N_0 is the noise power on RB c.

With all the settings above, our objective is to work out a smart joint CC selection and power control scheme to maximize the utilization of users' transmission power and mitigate the power offset effects brought by the nature of CA. An estimation method on the average number of RBs that a newly arrived UE can get from each CC is proposed to help the decision process, where UE location and current load CC conditions are carefully considered.

[1]There are several other varietal Shannon Formulas when the MAC and PHY mechanisms are specified, e.g., considering the HARQ constraints [92, 93]. But in this paper we consider the most basic one to give a clear demonstration on the relationship among rate, bandwidth and power.

4.2 Joint CC Selection and Power Control Algorithm

In this section, we first present the estimation method for average $N_{i,j}^{RB}$, leveraging the fairness properties of PF scheduling. Based on the estimation, the joint CC selection and power control scheme is put forward, considering user power constraints and offset effects. Pseudo-codes are provided at last to reveal the structure of the entire scheme at a glance.

4.2.1 Estimation Method for Average User Bandwidth

As mentioned in Subsection 6.1.2, PF method is adopted independently in each CC for Layer-2 packet scheduling. As shown in [90], PF scheduling guarantees that users within one CC can have nearly the same long-time average throughput (verified via simulation in Section 6.4). As the duration of one TTI is only $1ms$ [11], it is very reasonable that the users' inter-arrival time is much larger than one TTI, thus being sufficient for the users to have approximately equal throughput in one CC before next user arrival or departure.

Let $\hat{R}_{i,j}^{RB}$ denote the estimated average throughput of user i on each RB in CC j from the time when user i is assigned to CC j to that of next user arrival or departure in CC j. We call this period as the stable period of user i. The set of users transmitting on CC j within this period is denoted as $\mathbb{U}_j$. Thus

$$\hat{R}_{i,j}^{RB}\overline{N_{i,j}^{RB}} = A_j, \ \ \forall i \in U_j, \ j \in \{1, 2, ..., n\}, \tag{4.5}$$

where $\overline{N_{i,j}^{RB}}$ is the estimated average number of RBs in CC j allocated to user i within the stable period. A_j is a constant for each j. For CC j, denote the total number of RBs as $N_{RB,j}^{CC}$ and we have,

$$\sum_{i \in \mathbb{U}_j} \overline{N_{i,j}^{RB}} = N_{RB,j}^{CC}, \ \ \forall j \in \{1, 2, ..., n\}. \tag{4.6}$$

Combining (4.5) and (4.6), it can be found that if $\hat{R}_{i,j}^{RB}$ is known for each user i and CC j, then $\overline{N_{i,j}^{RB}}$ can be easily calculated. Recall the power control equation standardized in (4.1),

when the maximum transmission power is not exceeded, the equation can be rearranged as

$$\begin{aligned} &\tfrac{P_{i,j}{}'}{N_{i,j}^{RB}} = \Gamma \cdot 10^{\frac{\alpha L_i}{10}}\ (W), \\ &where \\ &\Gamma = 10^{\frac{(P_0-30)}{10}}, \\ &P'_{i,j} = 10^{\frac{(P_{i,j}-30)}{10}}. \end{aligned} \tag{4.7}$$

Γ here is a constant and $P'_{i,j}$ is another version of $P_{i,j}$ in the unit of W. As a result, the LHS of equation (4.7) indicates the average power from user i allocated to each RB in CC j. Denote the LHS of (4.7) as $\hat{P}_{i,c}$, turn it into the form of dBm and substitute it into (4.4). We can get $\hat{R}_{i,j}^{RB}$ as

$$\hat{R}_{i,j}^{RB} = W_{RB}\log_2\left[1 + \frac{\Gamma \cdot 10^{(\alpha-1)L_i/10}}{N_0}\right]. \tag{4.8}$$

Therefore, it can be concluded that if a newly arrived user i^* is assigned to CC j, combining (4.5) (4.6) and (4.8), we can calculate that the estimated average number of RBs that user i^* can occupy in CC j within its stable period is

$$\overline{N_{i^*,j}^{RB}} = \frac{N_{RB,j}^{CC}}{\hat{R}_{i^*,j}^{RB} \cdot \sum_{i\in\mathbb{U}_j}(\hat{R}_{i,j}^{RB})^{-1}}, \tag{4.9}$$

where $\hat{R}_{i^*,j}^{RB}$ and $\hat{R}_{i,j}^{RB}$ are achieved from equation (4.8). From equation (4.9) we can see that the proposed estimation method incorporates not only the user path loss (in the estimation of transmission rate per RB) but also the current load conditions (i.e., number of existing users in each CC) into consideration and thus being more adaptive and accurate.

4.2.2 Joint CC Selection and Power Control

Based on the above estimations and the standardized power control function in (4.1), the estimated total transmission power of UE i^* on set $\mathbb{C}_i$ (denoted as $\hat{P}_{i^*,total}$) is

$$\begin{aligned}
&\hat{P}_{i^*,total}\ \ (dBm)\\
&= \min\{P_{i^*,\max} - \hat{P}_{i^*,offset}, 10\log_{10}\left(\sum_{j\in\mathbb{C}_{i^*}} \hat{P}_{i^*,j}\right) - 30\},\\
&where\\
&\hat{P}_{i^*,j} = 10^{[P_0+10\log_{10}(\overline{N^{RB}_{i^*,j}})+\alpha L_{i^*}-30]/10}\ (W),\\
&\hat{P}_{i^*,offset} = (|\mathbb{C}_{i^*}| - 1 + \theta \sum_{j\in\mathbb{C}_{i^*}} \overline{N^{RB}_{i^*,j}})P_{back}\ (dB).
\end{aligned} \tag{4.10}$$

Note that $\hat{P}_{i^*,j}$ and $\hat{P}_{i^*,offset}$ are both estimated values achieved from equations (4.1) and (4.3), respectively. Then our estimation-based CC selection process can be described as that when a user i^* is admitted by the system, the CC selection entity will choose a CC subset $\mathbb{C}_{i^*}$ for user i^* that maximizes the estimated value $\hat{P}_{i^*,total}$ under the RR load balancing strategy, i.e.,

$$\begin{aligned}
&\max_{\mathbb{C}_i}\ \hat{P}_{i^*,total}\\
&s.t.\ Least - First\ load\ balancing;\\
&\quad Equation\ (4.9)\ and\ (4.10).
\end{aligned} \tag{4.11}$$

According to [11], currently CA only supports the aggregation of maximum 5 CCs. Therefore, an enumeration method is sufficient to find the optimal solution for (4.11).

Note that $\overline{N^{RB}_{i^*,j}}$ is only an estimated long-time average value used for decision making in the CC selection process. After the newly arrived user is assigned with a CC subset, the number of RBs that it can occupy is variant in different TTIs. Therefore, the power control function for each user must be dynamic and operate on the basis of actual RB-allocation circumstances per TTI. In one time slot, if the actual number of RBs user i can obtain from CC j is $N^{RB}_{i,j}$, then the total power user i is supposed to use (denoted as $P_{i,total}$) and the actual power offset $P_{i,offset}$ can be achieved via equation (4.1) and (4.3), respectively. Inspired

by [94], if the maximum available transmission power is exceeded after considering the effect of power offset, the user transmission power on each CC will be decreased by the factor $\Delta = (P_{i,total} - P_{i,max} + P_{i,offset})$. In this way, the actual power that user i spends on each CC j in $\mathbb{C}_i$ (referred to as $P_{i,j}$ in dBm) is shown below,

$$P_{i,j} = \begin{cases} P_0 + 10\log_{10}(N_{i,j}^{RB}) + \alpha L_i, & if\ \Delta \leq 0 \\ P_0 + 10\log_{10}(N_{i,j}^{RB}) + \alpha L_i - \Delta, & otherwise, \end{cases} \tag{4.12}$$

In each CC j, $P_{i,j}$ is equally shared by the $N_{i,j}^{RB}$ occupied RBs. Note that the subset $\mathbb{C}_i$ will not be changed once assigned to user i till the end of its transmission while $P_{i,j}$ will be dynamically adjusted every TTI.

To summarize the proposed joint CC selection and power control algorithm, a pseudo-code is presented in Algorithm 2. For a new user arrival, the total time complexity of the algorithm is calculated as,

$$O(nlog(n)) + O(n) = O(nlog(n)), \tag{4.13}$$

where n is the number of CCs. The first item on the LHS is the complexity for sorting the number of currently active users on each CC in an increasing order, which is in the order of $O(nlog(n))$. The second item on the LHS is the complexity for determining how many CCs should be assigned to the newly arrived user.

4.3 Performance Evaluation

4.3.1 Simulation Setup

To evaluate the performance of the proposed joint CC selection and power control strategy, system-level simulations are conducted in a single-cell SC-FDMA-based uplink scenario. Users arrive following a Poisson process with parameter λ and are uniformly distributed

Algorithm 1 Joint CC Selection and Power Control

Let N_j be the number of currently active users in CC j.
$N_j \leftarrow 0,\ \forall j \in \{1, 2, ..., n\}$; $t \leftarrow 0$
New user i^* arrives at tth TTI
while 1 **do**
    /*CC selection procedure*/
    Sort N_j in an increasing order. Let $Index$ contain the
  original indices corr. to the sorted list of N_j;
    $k \leftarrow 1$, $\hat{P}_{i^*,total} \leftarrow 0$, $\mathbb{C}_{i^*} \leftarrow \emptyset$;
    while $k \leq n$ **do**
        $j \leftarrow Index(k)$, $\mathbb{C}_{i^*} \leftarrow \mathbb{C}_{i^*} \cup \{j\}$;
        Calculate $\hat{P}_{i^*,total}$ and $\hat{P}_{i^*,offset}$ from equation
     (4.10);
        if $\hat{P}_{i^*,total} \leq 10^{(P_{i^*,max} - \hat{P}_{i^*,offset} - 30)/10}$ **then**
            $N_j \leftarrow N_j + 1$;
            $k \leftarrow k + 1$;
        else
            $\mathbb{C}_{i^*} \leftarrow \mathbb{C}_{i^*} \setminus \{j\}$;
            break;
        end if
    end while
    while No new user is admitted **do**
        /*Layer-2 PF scheduling*/
        for each $j \in \{1, 2, ..., n\}$ **do**
            PF scheduling;
        end for
        /*Power Control*/
        Determine $P_{i,j}$ for all users from equation (4.12)
        if User i finishes transmission **then**
            $N_j \leftarrow N_j - 1, \forall j \in \mathbb{C}_i$
        end if
        Proceed to next TTI, $t \leftarrow t + 1$
    end while
end while

within the cell coverage. The slow fading (distance-related path loss plus shadowing) remains unchanged for each user while frequency-selective fast fading is updated every TTI according to the Typical Urban (TU) channel model profile [95]. The shadowing effects are modeled as a normal variable with zero mean. Main default parameters and settings are

summarized in Table 5.1 for reference.

In addition, to better illustrate the performance gain, the path-loss-threshold-based CC assignment algorithm in [50] is also simulated for comparison. The algorithm derives a path-loss threshold to classify users into power-constrained and non-power-constrained groups, and assign all CCs to the former but only one CC to the latter.

Table 4.1 Main Default Simulation Parameters

Parameters	Values
User Arrival Rate , λ	1/50 ms^{-1}
User Payload P_i	$15Mbits$
Cell Radius	$1500m$
Path Loss Factor β	3
Shadowing Statistics	$\mu = 0\ dB$, $\sigma = 8\ dB$
Noise Power per RB, N_0	$-116dBm$
$[\alpha,\ P_0]$	$[0.6, -58dBm]$
Max Tx Power per UE, $P_{i,max}$	$23dBm$
Power Offset Constant, $[P_{back},\ \theta]$	$[3dB,\ 0.01]$
Number of Carriers, n	5
Number of RBs per CC, $N_{RB,j}^{CC}$	50
Bandwidth per RB, W_{RB}	180KHz
TD Scheduling	Least-First
FD Scheduling	Proportional Fair

4.3.2 Simulation Results

We first simulate the user fairness performance of PF scheduling to validate the proposed estimation method. The fairness metric used in [90] is adopted in our verification, which is a data-rate fairness criterion expressed as:

$$F(\Delta t) = [\sum_{i=1}^{N} R_i(\Delta t)]^2 / [N \cdot \sum_{i=1}^{N} {R_i}^2(\Delta t)], \tag{4.14}$$

where $R_i(\Delta t)$ is the actual data rate that user i achieved in Δt when N users are sharing RBs in one CC. It can be seen that $F(\Delta t)$ reaches its maximum value 1 only when all the users have equal actual data rates in Δt. Simulations are conducted in one single CC for different Δt and N, and the results are shown in Fig. 4.1. It can be observed that for all simulated N,

the user fairness metric converges to 1 eventually after a sufficient time duration. The reason is that PF scheduling considers not only the estimated instantaneous but also the average past user throughput and makes a fair tradeoff between current channel conditions and user throughput history. Besides, the figure shows that the convergence speed is smaller with larger N, which matches the intuition well that the more users in one CC, the longer time it takes for the PF scheduler to balance all the users. Note that since one TTI is only $1ms$, the user inter-arrival time is sufficiently long to guarantee a good fairness performance in most cases.

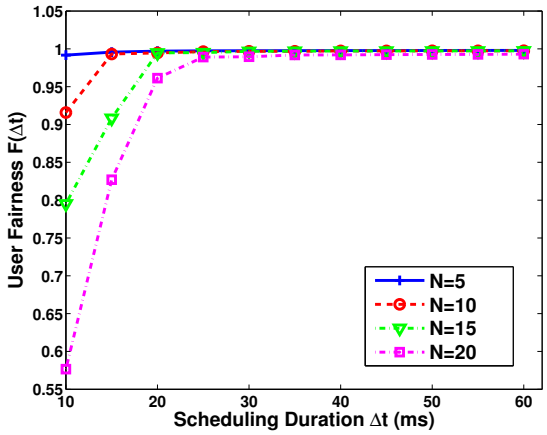


Figure 4.1. User fairness under different values of Δt and N

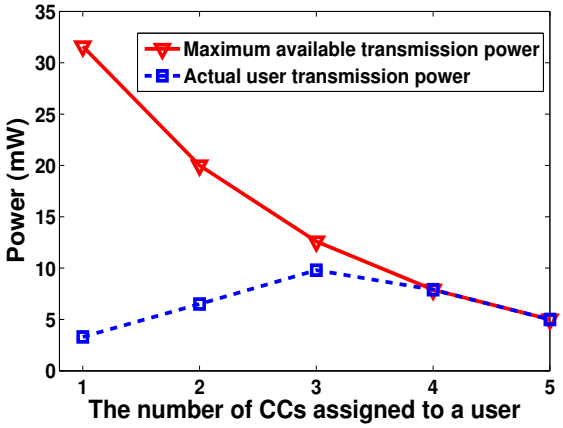


Figure 4.2. Power utilization with the number of assigned CCs.

Then, we examine how the power utilization of the user terminal changes with the number of CCs assigned to one user. The results are shown in Fig. 4.2. It can be seen that with the increase of the number of assigned CCs, the maximum available transmission power in the user terminal keeps increasing due to the power offset effects; the actual transmission power that the terminal uses first increases and then reduces. The highest power utilization (i.e., the user actual transmission power over the total terminal power) is achieved when the user is assigned with 3 CCs instead of all 5 CCs. The observation validates that it is not always good to assign a user with as many CCs as possible.

At last, we compare the performance between our proposed estimation-based CC se-

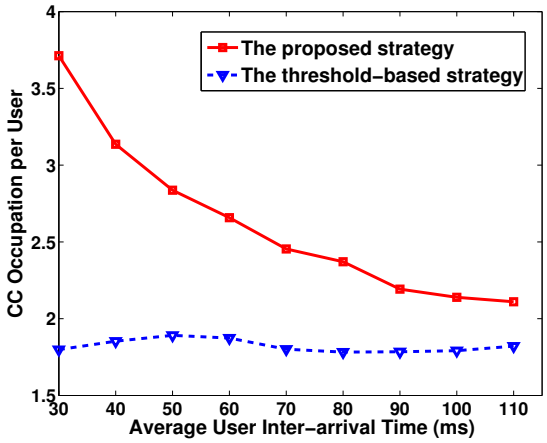


Figure 4.3. CC occupation per user vs. average user inter-arrival time

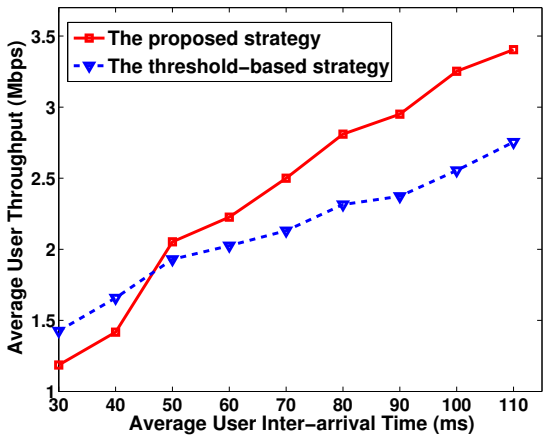


Figure 4.4. CC occupation per user vs. average user inter-arrival time

lection strategy and the threshold-based one. Two measurements are emphasized in our simulation, i.e., CC occupation per user and the average user throughput. The results under different user inter-arrival times (i.e., $1/\lambda$) are shown in Fig. 4.3 and Fig. 4.4, respectively.

CC occupation indicates the average number of CCs each user can be assigned in the whole simulation process. From Fig. 4.3, it can be seen that the CC occupation per user under our proposed strategy is higher than that under the threshold-based one. Besides, the former one decreases when the average user inter-arrival time increases while the latter one remains almost unchanged. The reason is that under the threshold-based algorithm, the number of CCs one user can get is only related to its pass loss. In such a case, when one user arrives, the probability whether it is assigned with one CC or all CCs is a constant regardless of the load conditions in each CC. As a result, the average CC occupation does not change. However, with our proposed strategy, when the average inter-arrival time $1/\lambda$ increases, it is more likely that more users are active in one CC, resulting in a decrease in the average number of RBs one user can get from one CC. In this case, one user will spend less power on one CC and thus can afford concurrent transmissions in more CCs. In other words, higher CC load conditions can make more previously power-constrained users become non-power-constrained, and thus being assigned with more CCs. By taking into account the time-varying

CC load conditions, the CC occupation under the proposed strategy is relatively larger than that in the compared strategy.

The throughput comparison is shown in Fig. 4.4. It can be observed that for most $1/\lambda$ values, the average user throughput under the proposed strategy is considerably higher than that achieved under the estimation-based strategy, however, when the CCs are heavily loaded (i.e., when $1/\lambda$ is very small, e.g., 30 and 40), the results are opposite even though the CC occupation per user under the proposed strategy is very high. The reason for this interesting phenomenon is that when the CCs are heavily loaded, the newly admitted user i has a higher probability to be assigned with multiple CCs since the power it needs to use on each CC is smaller. In this way, although the throughput of user i is improved, the throughput of other existing users in CCs belonging to $\mathbb{C}_i$ will be affected. As the size of $\mathbb{C}_i$ is very likely bigger under the proposed strategy, more existing users will be affected, which will counterbalance the throughput gain achieved by user i and even result in a worse average user throughput. But when the average user inter-arrival time increases, the throughput gain surpasses the loss and the advantage of our proposed strategy shows up.

Another key reason for the throughput gain is that the infamous power offset effect is properly dealt with in the proposed strategy. Since the maximum available transmission power will be reduced with more CCs, instead of assigning all the CCs to the non-power-constrained users, a subset of CCs is carefully chosen for each newly admitted non-power-constrained user based on the current CC load conditions so that users' actual transmission power can be maximized.

4.4 Summary

In this chapter, we have studied the cross-layer RRM performance of uplink CA in LTE-A systems. A joint CC selection and power control strategy is proposed to enhance the average user throughput, considering the user power constraints and offset effects. In specific, an estimation-based method is first put forward to calculate the expected number of RBs that

one newly admitted user can get from each CC, with considering the dynamic CC load conditions. Then a user-power-utilization maximization problem is formulated to determine the optimal CC subset. Dynamic power control are conducted thereafter in every TTI according to the actual number of occupied RBs for each user. Extensive simulation results have demonstrated that the average user throughput under the proposed strategy is considerably higher due to better power utilization. For the future work, we will investigate the impact of intercell interference and different channel model profiles on the CC selection performance with different QoS metrics.

Chapter 5

Probabilistic QoS Provisioning in LTE-A HetNets with Partial Spectrum Usage

In this chapter, we investigate the QoS provisioning issue in LTE-A HetNets with PSU mechanism. Users that require ubiquitous mobility support or low-rate services connect to MBSs, while FBSs are deployed to serve the users that require high-data-rate transmissions. Hybrid-access FCells are considered where a subset of carriers are reserved for the FCell subscribers (FSs), while another disjoint subset is open to provide paid services to FCell nonsubscribers (FNSs). Two challenging issues are addressed: i) under PSU, both the FCell random behaviors and the inter-macro interference are deliberately modeled and incorporated into the performance analysis framework, with considering the users' QoS requirements and CA capabilities; and ii) the interplay between MBSs and FBSs is formulated into correlated utility maximization problems to determine the optimal RA decisions.

Specifically, we first model the locational randomness of MBSs, FBSs and users into Poisson Point Processes (PPPes) [79]. Stochastic Geometry theory is exploited to obtain the signal-to-interference-plus-noise (SINR) distributions and ergodic throughput (measures long-term average user throughput) for different user types in each CC. The deduction considers PSU mechanism, user CA capabilities and configurable user bandwidth. To satisfy the QoS requirements of different user types with appropriate bandwidth assignment, the

concept *effective bandwidth* [78] is leveraged to provide a unified bandwidth for each user type based on the derived SINR distributions. With the derived effective bandwidth, users are provided with probabilistic QoS guarantee. Particularly, to make the decision process for effective bandwidth practical and tractable, an heuristic iterative algorithm named QA-EB algorithm is proposed. Then, the interplay of RA between MCells and FCells is formulated into a two-level Stackelberg game. In the game, based on the aggregate throughput of FBSs, MBSs first impose an interference-related price upon FBSs, and FBSs adjust their PSU policy accordingly. A backward induction method is proposed to achieve the Stackelberg equilibrium (i.e., optimal price and PSU policy). The method shows how the price and PSU policy are tuned to maximize the utilities of both parties. Finally, simulation results validate our analytical ones, and the Stackelberg equilibrium is demonstrated under different user QoS requirements and CA capabilities.

5.1 System Model

In this section, the PPP-based HetNets layout is first presented. The bandwidth access mechanisms and physical channel model are then introduced, followed by the interaction model between macro and femto cells.

5.1.1 Network Deployment

As shown in Fig. 5.1, we consider an arbitrary region A with area $|A|$, where the MBSs and FBSs are deployed as homogeneous PPPes with density measure λ^{MBS} and λ^{FBS}, respectively. In other words, MBSs (or FBSs) are uniformly distributed within A with the total number following a Poisson distribution - $Poisson(\lambda^{MBS}|A|)$ (or $Poisson(\lambda^{FBS}|A|)$). Due to much smaller transmission powers, FCell coverage is much smaller than MCell coverage. The sets of MBSs and FBSs are denoted as Φ^{MBS} and Φ^{FBS}, respectively.

The MCell users (MUs) are distributed within A following a homogeneous PPP with density λ^{MU}. Each MU connects to its nearest MBS for service. Under such an association

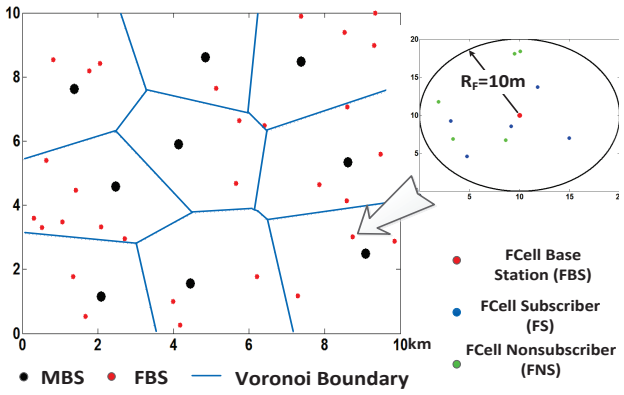


Figure 5.1. The network layout of HetNets.

policy, the actual coverage of an MBS becomes a Voronoi cell [80] where any point in a Voronoi cell has a shorter distance to the associated MBS than to other MBSs. The FSs (or FNSs) are distributed as a homogeneous PPP with density λ^{FS} (or λ^{FNS}), in a disk coverage of FCell with radius R_F.

5.1.2 Bandwidth Allocation Mechanisms

The system bandwidth consists of N carriers. Each carrier i ($\in \{1, ..., N\}$) is further divided into P_i orthogonal Physical Resource Blocks (PRBs[1]), each with bandwidth W_{PRB}. A PRB is the minimum bandwidth allocation unit in LTE-A systems, as shown in Fig. 5.2. In carrier i, the PRBs are all orthogonal. One PRB can only be assigned to one user within 1 subframe of 1ms, while one user can occupy several PRBs concurrently. The PRBs assigned to one user can be contiguous or not [11].

All users are assumed to have a CA capability n_{agg} indicating that a user can transmit on n_{agg} carriers simultaneously. One MU requires a minimum throughput r_u^{MU}, while one FS or FNS enjoying high-speed services requires minimum throughput r_u^F ($r_u^F > r_u^{MU}$). To provide the users with probabilistic guarantee on the throughput requirement, each MU, FS and FNS is assigned with effective bandwidth W_i^{MU}, W_i^{FS} and W_i^{FNS}, respectively in

[1]LTE-A is built upon the Orthogonal Frequency Division Multiple Access (OFDMA) technology, and a PRB consists of 12 contiguous OFDMA subcarriers.

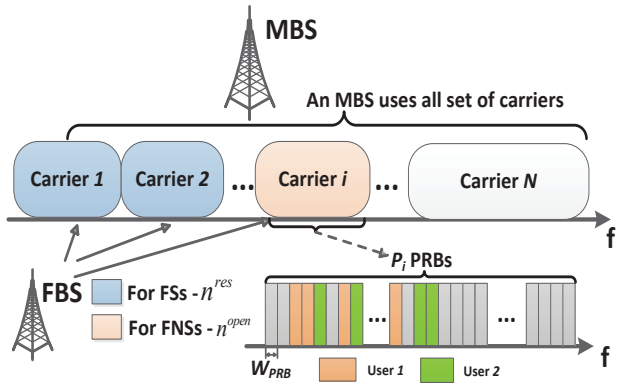


Figure 5.2. Bandwidth structure of HetNets under PSU.

carrier i, such that

$$\begin{aligned} &\Pr\left(\sum_{i=1}^{N} R_i^{MU}(W_i^{MU}) < r_u^{MU}\right) < e \ll 1, \\ &\Pr\left(\sum_{i=1}^{N} R_i^{T}(W_i^{T}) < r_u^{F}\right) < e \ll 1,\ T \in \{FS, FNS\}, \end{aligned} \tag{5.1}$$

where $R_i^T(W_i^T)$ denotes the ergodic throughput that a type-T user can get from carrier i given effective bandwidth W_i^T, and e denotes a small positive value much smaller than 1. Eq. (6.5) means that the total ergodic rate of all carriers for a user should be smaller than its required throughput with a very small probability.

For the access mechanism, each MBS operates on all N carriers to serve MUs, while each hybrid-access FBS randomly and independently chooses n^{res} carriers to serve the FSs and n^{open} disjoint carriers for open access, satisfying $n^{res} + n^{open} \leq N$. For one MBS or FBS, denote the number of type-T users choosing carrier i for transmission as N_i^T and the maximum number of users that can be concurrently served in carrier i as $N_i^{T,ser} = \left\lfloor P_i \cdot W_{PRB}/W_i^T \right\rfloor,\ T \in \{MU, FS, FNS\}$. In each subframe, if $N_i^T \leq N_i^{T,ser}$, the system will randomly choose N_i^T pieces of W_i^T bandwidth for type-T users; otherwise, time-sharing scheduling is adopted to randomly select $N_i^{T,ser}$ type-T users to transmit. In this way, each type-T user can be served with equal long-term time-proportion within one carrier.

5.1.3 Physical Channel Model

The path loss and fast fading effects are considered in this work. The shadowing effects are not included as [83] has proved that the shadowing can be well approximated by the randomness of the Poisson distributed BS locations. This is a strong justification that the distribution of MBSs can be modelled as a PPP.

We consider that the power spectrum densities (PSD) of MBSs and FBSs are fixed in carrier i and denoted as P_i^{MBS} and P_i^{FBS}, respectively. For a user, its received PSD in carrier i from an MBS (or FBS) B with a distance of D_B is

$$P_i^r = P_i^T H {D_B}^{-\alpha_i},\ B \in \Phi^T,\ T \in \{MBS, FBS\} \tag{5.2}$$

where H is the fast fading channel gain and α_i is the path loss exponent. The fast fading of the useful signal is considered as Rayleigh fading, so the fast fading channel gain follows an exponential probability density function[2] (pdf), i.e., Exp(μ). For simplicity, we set μ as 1. The fast fading of the interference signals is considered as generally distributed.

5.1.4 Economic Interaction between Macro and Femto Cells

The objective of both parties is to maximize their own utilities, which are expressed as the weighted summations of multiple parts of profits. Each MBS charges MUs' services with unit price g^{MU}/bit. Meanwhile, to preserve the MU performance from the interference by FBSs, MBSs impose an interference unit price y_i over FBSs for interfering the carrier i. For analytical simplicity, y_i's are set to be equal for all carriers and denoted as y in the rest of the chapter. Besides, an upper bound y^{max} is imposed on y to avoid overcharging, which is reasonable in practical. Therefore, there are two parts of profits for one MBS: the profits from MU services and the profits from charging all FBSs within its coverage.

Bearing the interference price y, each FBS will optimize the subsets of carriers assigned

[2]If a random variable Ra is Rayleigh distributed, then its power Ra^2 is exponentially distributed with parameter μ.

to FSs and FNSs, i.e., n^{res} and n^{open}, considering the effective bandwidth of all the user types and CA capabilities. One FBS pays unit price g^{FS}/bit for the FSs' services and can gain profits with unit price g^{FNS}/bit from FNSs. Therefore, the total utility of one FBS is the three-fold: profits from FNSs, service payment for FSs, and the interference cost charged by the MBSs.

5.2 Probabilistic Analysis on User Performance for HetNets with PSU

In this section, the Stochastic Geometry is exploited to model the HetNets interference with PSU, considering user QoS requirements and CA capabilities. Both the FCell randomness and the multi-macro interference are included. Specifically, the SINR distributions and ergodic rates for each type of users in each carrier are first derived (Subsection 5.2.1 and 5.2.2), and then the effective bandwidth is finalized according to the user QoS requirement (Subsection 5.2.3).

5.2.1 SINR Distributions and User Ergodic Rates

The SINR distribution of an MU in carrier i is derived first. The probability that $SINR_i^{MU}$ is larger than a threshold β is

$$\begin{aligned}
&\mathrm{P}(SINR_i^{MU} > \beta) = \mathrm{P}(\frac{P_i^{MBS} H {D_{B_0}}^{-\alpha_i}}{I_i^{MBS}+I_i^{FBS}+n_0} > \beta) \\
&= \mathrm{P}(H > \frac{\beta(I_i^{MBS}+I_i^{FBS}+n_0){D_{B_0}}^{\alpha_i}}{P_i^{MBS}}), \\
&where\ I_i^{MBS} = \sum_{B\in\Phi_i^{MBS}\setminus B_0} P_i^{MBS} H_i^{MBS} {D_B}^{-\alpha_i} \\
&\quad I_i^{FBS} = \sum_{F\in\Phi_i^{FBS}} P_i^{FBS} H_i^{FBS} {D_F}^{-\alpha_i}.
\end{aligned} \tag{5.3}$$

In Eq. (6.11), B_0 is the associated MBS of the considered MU. Notation Φ_i^{MBS} (Φ_i^{FBS}) denotes the set of MBSs (FBSs) that use the same PRBs with the considered MU in carrier i; I_i^{MBS} (I_i^{MBS}) denotes the interference PSD from Φ_i^{MBS} (Φ_i^{FBS}); and H_i^{MBS} (H_i^{FBS}) denotes the fast fading channel gain between the considered user to the MBSs (FBSs). As $H \sim \text{Exp}(1)$, we have $\text{P}(H > h) = e^{-h}$. Then

$$\begin{aligned}&\text{P}(H > \tfrac{(I_i^{MBS}+I_i^{FBS}+n_0){D_{B_0}}^{\alpha_i}\beta}{P_i^{MBS}})\\ &= E_{\Phi_i^{MBS}\setminus B_0,\Phi_i^{FBS},H_i^{MBS},H_i^{FBS},D_{B_0}}\\ &\qquad [\exp(-\tfrac{(I_i^{MBS}+I_i^{FBS}+n_0)D_{B_0}^{-\alpha_i}\beta}{P_i^{MBS}})],\end{aligned} \tag{5.4}$$

where $E[\cdot]$ denotes the expectation and $\Phi_i^{MBS}\setminus B_0$ is the set Φ_i^{MBS} excluding MBS B_0. As randomness exists in $\Phi_i^{MBS}\setminus B_0$, Φ_i^{FBS}, H_i^{MBS}, H_i^{FBS} and D_{B_0}, $\text{P}(SINR_i^{MU} > \beta)$ should be an expectation over all these items. Proposition 1 gives the derived SINR distribution of one MU in carrier i.

Proposition 1: In the HetNets described in the system model, given the effective bandwidth of all the user types (i.e., W_i^{MU}, W_i^{FS}, and W_i^{FNS}), the probability that the SINR of one MU in carrier i is larger than a threshold β is given as

$$\begin{aligned}&\text{P}(SINR_i^{MU} > \beta)\\ &= \textstyle\int_0^{+\infty} 2\pi\lambda^{MBS} d e^{-\pi\lambda^{MBS}d^2} e^{-n_0 d^{\alpha_i}\beta/P_i^{MBS}}\\ &\quad \cdot \exp\{-2\pi\lambda_i^{MBS}\eta(d, H_i^{MBS}, \beta)\}\\ &\quad \cdot \exp\{-2\pi\theta_i^{FBS}\lambda^{FBS,usa}\epsilon(d, H_i^{FBS}, \beta, A)\}d(d),\\ &\text{where}\\ &\eta(d, H_i^{MBS}, \beta) = -\tfrac{1}{2}d^2 + \tfrac{1}{2}d^2 E_{H_i^{MBS}}\{e^{-\beta H_i^{MBS}} +\\ &\qquad (\beta H_i^{MBS})^{2/\alpha_i}\left[\Gamma(1-\tfrac{2}{\alpha_i}, 0) - \Gamma(1-\tfrac{2}{\alpha_i}, \beta H_i^{MBS})\right]\},\\ &\epsilon(d, H_i^{FBS}, \beta, A) = \tfrac{1}{2}d^2\Gamma(1-\tfrac{2}{\alpha_i}, 0)E_{H_i^{FBS}}\left[(A\beta H_i^{FBS})^{2/\alpha_i}\right],\\ &\Gamma(s,t) = \textstyle\int_t^{+\infty} x^{s-1}e^{-x}dx, \text{ and } A = P_i^{FBS}/P_i^{MBS}.\end{aligned} \tag{5.5}$$

Please refer to [65] for the detailed derivations of Proposition 1. One interesting observation in Proposition 1 is that the SINR distribution is not related to the effective bandwidth of any user type. This is because fixed transmission PSDs are considered for both MBSs and FBSs in the system model. Therefore, the ergodic throughput of the MUs in carrier i can be calculated as

$$\begin{aligned} R_i^{MU} &= E_{SINR_i^{MU}}\left[\mathrm{Q}_{s|i}^{MU} W_i^{MU} \log(1+SINR_i^{MU})\right] \\ &= \frac{\mathrm{Q}_{s|i}^{MU} W_i^{MU}}{ln2} \int_0^{+\infty} \mathrm{P}(\ln(1+SINR_i^{MU}) > t) dt \\ &\overset{\beta=e^t-1}{=} \frac{\mathrm{Q}_{s|i}^{MU} W_i^{MU}}{ln2} \int_0^{+\infty} \frac{1}{1+\beta} \mathrm{P}(SINR_i^{MU} > \beta) d\beta, \end{aligned} \tag{5.6}$$

where $\mathrm{Q}_{s|i}^{MU}$ denotes the service probability that an MU can be scheduled to have W_i^{MU} bandwidth conditioning on that it selects carrier i. Variables θ_i^{MBS}, θ_i^{FBS} and $\mathrm{Q}_{s|i}^{MU}$ are closely related to the PSU policy and calculated in Subsection 5.2.2.

Similarly as MUs, the SINR distribution of FSs is given in Proposition 2.

Proposition 2:In the HetNets described in the system model, the probability that the SINR of one FS in carrier i is larger than a threshold β can be expressed as

$$\begin{aligned} &\mathrm{P}(SINR_i^{FS} > \beta) = \int_0^{R_F} \frac{2d}{R_F^2} e^{-n_0 d^{\alpha_i} \beta / P_i^{FBS}} F d(d) \\ &where \;\; F = \exp\{-2\pi\theta_i^{FBS,usa} \lambda^{FBS} \tau(d, H_i^{FBS}, \beta)\} \\ &\qquad \cdot \exp\{-2\pi\theta_i^{MBS,usa} \lambda^{MBS} \rho(d, H_i^{MBS}, \beta, B)\}, \\ &\tau(d, H_i^{FBS}, \beta) = \frac{d^2}{2}\Gamma(1-\frac{2}{\alpha_i}) E_{H_i^{FBS}}\left[(\beta H_i^{FBS})^{\frac{2}{\alpha_i}}\right], \\ &\rho(d, H_i^{MBS}, \beta, B) = \tau(d, B H_i^{MBS}, \beta), \\ &B = P_i^{MBS}/P_i^{FBS}. \end{aligned} \tag{5.7}$$

Please refer to [65] for the detailed derivations of Proposition 2. Therefore, the ergodic throughput of FSs in carrier i, R_i^{FS}, is calculated similarly as Eq. (5.6),

$$R_i^{FS} = \frac{\mathrm{Q}_{s|i}^{FS} W_i^{FS}}{ln2} \int_0^{+\infty} \frac{1}{1+\beta} \mathrm{P}(SINR_i^{FS} > \beta) d\beta, \tag{5.8}$$

where $\mathrm{Q}_{s|i}^{FS}$ denotes the user service probability of one FS in carrier i given that it uses carrier

i. Following the same procedure, the SINR distribution of FNS in carrier i, $P(SINR_i^{FNS} > \beta)$, and ergodic throughput R_i^{FNS} can be calculated similarly as Eq. (5.7) and (5.8) where the superscript *"FS"* is replaced with *"FNS"*.

5.2.2 User Service Probability and Bandwidth Usage Probability

In this subsection, we calculate the user service probabilities ($\mathrm{Q}_{s|i}^{MU}$, $\mathrm{Q}_{s|i}^{FS}$ and $\mathrm{Q}_{s|i}^{FNS}$) that one user can be served by an MBS or FBS, and the bandwidth usage probability ($\theta_i^{MBS,usa}$ and $\theta_i^{FBS,usa}$) that a portion of effective bandwidth in one carrier is occupied by any user in one MBS or FBS. All the probabilities are closely related to the PSU policy (n^{res} and n^{open}) and the user CA capabilities (n_{agg}) and are conditioned on that the user selects carrier i to transmit.

To calculate the MBS-related probabilities $\mathrm{Q}_{s|i}^{MU}$ and $\theta_i^{MBS,usa}$, the number of MUs in one MCell needs to be calculated first. Denote the number of MUs in a Voronoi cell and the cell size as N^{MU} and S, respectively. As the MUs are distributed as a PPP with density λ^{MU}, the number of MUs in one Voronoi cell with area S (denoted as N^{MU}) follows Poisson($\lambda^{MU}S$). Thus, we have

$$\mathrm{P}(N^{MU} = k|S) = \frac{(\lambda^{MU}S)^k e^{-\lambda^{MU}S}}{k!},\ k = 0, 1, \ldots \tag{5.9}$$

And

$$\mathrm{P}(N^{MU} = k) = \int_0^{+\infty} P(N^{MU} = k|S) f(S) dS, \tag{5.10}$$

where $f(S)$ is the pdf of S. As indicated in [87], a simple but accurate enough approximation of $f(S)$ is given as,

$$f(S) = \frac{343}{15}\sqrt{\frac{7}{2\pi}}(S\lambda^{MBS})^{\frac{5}{2}} e^{-\frac{7}{2}S\lambda^{MBS}} \lambda^{MBS}. \tag{5.11}$$

By substituting Eq. (5.11) into Eq. (5.10), the distribution of N^{MU} can be obtained. As one MU can access any carrier and concurrently transmit on n_{agg} carriers, the probability that one MU chooses carrier i is n_{agg}/N. Then the probability that there are totally k MUs in one

cell among which l MUs choose carrier i is denoted as $P^{MU}_{l,k|i}$ and calculated as

$$\mathrm{P}^{MU}_{l,k|i} = C^l_k(\frac{n_{agg}}{N})^l(1-\frac{n_{agg}}{N})^{k-l}\mathrm{P}(N^{MU}=k). \tag{5.12}$$

Given l MUs choose carrier i, the probability that one MU can have bandwidth from carrier i is

$$\min\{1, \frac{P_i W_{PRB}/W_i^{MU}}{l}\} = \min\{1, \frac{P_i W_{PRB}}{W_i^{MU} l}\}. \tag{5.13}$$

Then user service probability $\mathrm{Q}^{MU}_{s|i}$ can be achieved by averaging Eq. (5.13) over $\mathrm{P}^{MU}_{l,k|i}$,

$$\mathrm{Q}^{MU}_{s|i} = \sum_{k=1}^{\infty}\sum_{l=1}^{k} \min\{1, \frac{P_i W_{PRB}}{W_i^{MU} l}\} P^{MU}_{l,k|i}. \tag{5.14}$$

Similarly, the bandwidth usage probability $\theta_i^{MBS,usa}$ is calculated as

$$\theta_i^{MBS,usa} = \sum_{k=1}^{\infty}\sum_{l=1}^{k} \min\{1, \frac{W_i^{MU} l}{P_i W_{PRB}}\} P^{MU}_{l,k|i}. \tag{5.15}$$

Then, the FBS-related probabilities (i.e., $\mathrm{Q}^{FS}_{s|i}$, $\mathrm{Q}^{FNS}_{s|i}$ and $\theta_i^{FBS,usa}$) can be calculated. As the area of one FBS coverage is πR_F^2, the total number of FSs in one FCell (denoted as N^{FS}) is Poisson distributed with $\lambda^{FS}\pi R_F^2$. Thus, we have

$$\mathrm{P}(N^{FS}=k) = \frac{(\lambda^{FS}\pi R_F^2)^k}{k!} e^{-\lambda^{FS}\pi R_F^2}. \tag{5.16}$$

Similarly as the calculation of $Q^{MU}_{s|i}$, $\mathrm{Q}^{FS}_{s|i}$ is calculated as

$$Q^{FS}_{s|i} = \begin{cases} \sum\limits_{k=1}^{\infty} \min\{1, \frac{P_i W_{PRB}}{k W_i^{FS}}\} P(N^{FS}=k), & if\ n_{agg} \geq n^{res} \\ \sum\limits_{k=1}^{\infty}\sum\limits_{l=1}^{k} \min\{1, \frac{P_i W_{PRB}}{l W_i^{FS}}\} P^{FS}_{l,k|i}, & otherwise, \end{cases} \tag{5.17}$$

$$where\ \mathrm{P}^{FS}_{l,k|i} = C^l_k(\tfrac{n_{agg}}{n^{res}})^l(1-\tfrac{n_{agg}}{n^{res}})^{k-l}\mathrm{P}(N^{FS}=k).$$

The probability $Q_{s|i}^{FNS}$ is calculated exactly the same way as that of $Q_{s|i}^{FS}$. For $\theta_i^{FBS,usa}$, the probabilities that carrier i is assigned to FSs and FNSs are n^{res}/N and n^{open}/N, respectively. For either possibility, the bandwidth usage probability is calculated similarly with Eq. (5.15), i.e.,

$$\begin{aligned}
&\theta_i^{FBS} = \tfrac{n^{res}}{N} P_1 + \tfrac{n^{open}}{N} P_2, \\
&where \\
&P_1 = \\
&\begin{cases} \sum\limits_{k=1}^{\infty} \min\{1, \frac{kW_i^{FS}}{P_i W_{PRB}}\} P(N^{FS} = k), & if\ n_{agg} \geq n^{res} \\ \sum\limits_{k=1}^{\infty} \sum\limits_{l=1}^{k} \min\{1, \frac{lW_i^{FS}}{P_i W_{PRB}}\} P_{l,k|i}^{FS\prime}, & otherwise, \end{cases}
\end{aligned} \tag{5.18}$$

The probability P_2 is FNS-related and can be calculated similarly with P_1.

5.2.3 QoS-Aware Effective Bandwidth: Formulation and Algorithm

The effective bandwidth for each type of user is finalized based on the derived user SINR distributions. For analytical simplicity, the intra-band contiguous CA [11] is considered where the radio characteristics of all carriers are the same, so each carrier contribute equal portion of throughput for each user. Thus for MUs, the minimum throughput requirement on carrier i is r_u^{MU}/n_{agg}. According to Eq. (6.5), W_i^{MU} should be determined such that

$$\mathrm{P}(Q_{s|i}^{MU} W_i^{MU} \log(1 + SINR_i^{MU}) < r_u^{MU}/n_{agg}) < e \ll 1, \tag{5.19}$$

which can be rearranged as

$$\mathrm{P}(SINR_i^{MU} < 2^{r_u^{MU}/(n_{agg} Q_{s|i}^{MU} W_i^{MU})} - 1) < e. \tag{5.20}$$

If β is equal to $2^{r_u^{MU}/(n^{agg} Q_{s|i}^{MU} W_i^{MU})} - 1$, Eq. (5.5) can be leveraged to achieve the value range of W_i^{MU}. As one PRB is the minimum bandwidth allocation unit in LTE-A systems, W_i^{MU} should be an integral multiple of W_{PRB}. Then W_i^{MU} can be finalized as the product of W_{PRB} and an integer value denoted as m_i^{MU}. The physical meaning of m_i^{MU} is the minimum

number of PRBs in carrier i that can satisfy Eq. (5.20), i.e.,

$$\begin{aligned} & m_i^{MU} = \min_{m} \ m \\ & s.t. \ W_i^{MU} = mW_{PRB}; \\ & \quad 0 \leq m \leq P_i, \ m \in \mathbb{Z}^+; \\ & \quad Eq. \ (5.20). \end{aligned} \tag{5.21}$$

According to Eq. (5.20), W_i^{MU} is closely related to $Q_{s|i}^{MU}$ and $\theta_i^{MBS,usa}$ which are further determined by the PSU policy and CA capabilities. Therefore, W_i^{MU} is jointly determined by the PSU policy, user QoS requirements and CA capabilities. For the integer values of FSs (or FNSs), denoted as m_i^{FS} (or m_i^{FNS}), the derivation is the same as that of MUs except the minimum throughput requirements in carrier i, which is $r_u^F / \min\{n^{res}, n_{agg}\}$ (or $r_u^F / \min\{n^{open}, n_{agg}\}$).

It can be seen from Eq. (5.21) that for any type of users, the optimization problem to calculate the effective bandwidth is constrained integer non-convex. In addition, the determination of its effective bandwidth is highly dependent on the effective bandwidth of the other types, which is because the constraint on its SINR distribution (i.e., the third constraint of the optimization problem) is closely related to the effective bandwidth of the other types. Therefore, it is infeasible to obtain the optimum in polynomial time.

To make the proposed strategy tractable and practical, a heuristic algorithm, referred to as QoS-aware effective bandwidth (QA-EB) algorithm in this chapter, is proposed. The basic idea of the QA-EB algorithm is to augment m_i^{MU}, m_i^{FS} and m^{FNS} step by step according to a specified priority. Each of the above three variables starts from 1 with augmentation step 1. Each time when one variable increases 1, the algorithm checks whether the SINR constraints of the user types with higher priority are satisfied. If the constraints are satisfied, the variable of the user type with the next lower priority is augmented; otherwise the variable of the user type with the highest priority and unsatisfied SINR constraint is augmented. The algorithm stops when the SINR constraints of all the user types are satisfied. Note that The priority of user types can be determined according to the vendor/operator's preference, and different

priority assignments may lead to different effective bandwidth sets $\{m_i^{MU}, m_i^{FS}, m_i^{FNS}\}$. In this work, the MUs and FNSs are given the highest and lowest priority, respectively. The reason is that the MUs are more sensitive to the change of effective bandwidth of FSs and FNSs, which has been validated through simulations in Section 6.4.

Algorithm 2 QA-EB Algorithm

/* **Initialization** */
Define auxiliary variables
$A_1 := r_u^{MU}/(n_{agg}Q_{s|i}^{MU}W_i^{MU})$,
$A_2 := r_u^{FS}/(\min\{n_{agg}, n^{res}\}Q_{s|i}^{FS}W_i^{FS})$,
$A_3 := r_u^{FNS}/(\min\{n_{agg}, n^{open}\}Q_{s|i}^{FNS}W_i^{FNS})$;
Define events E_1, E_2, E_3 as
$E_1 := \mathrm{P}(SINR_i^{MU} < 2^{A_1} - 1) < e$,
$E_2 := \mathrm{P}(SINR_i^{FS} < 2^{A_2} - 1) < e$,
$E_3 := \mathrm{P}(SINR_i^{FNS} < 2^{A_3} - 1) < e$;
$m_i^{MU} \leftarrow 1, m_i^{FS} \leftarrow 1, m_i^{FNS} \leftarrow 1$;
/***Loop Augmentation***/
EndFlag←FALSE;
while EndFlag == FALSE **do**
    while E_1 is TRUE and EndFlag == FALSE **do**
        if E_2 is FALSE **then**
            $m_i^{FS} \leftarrow m_i^{FS} + 1$;
        end if
        while E_1 is TRUE and E_2 is TRUE and EndFlag == FALSE **do**
            if E_3 is TRUE **then**
                EndFlag ← TRUE;
            else
                $m_i^{FNS} \leftarrow m_i^{FNS} + 1$;
            end if
        end while
    end while
    if E_1 is FALSE **then**
        $m_i^{MU} \leftarrow m_i^{MU} + 1$;
    end if
end while

Remark: According to Eq. 6.5, if the throughput requirement for one user (i.e., r_u^{MU} and r_u^F) is higher, the effective bandwidth (i.e., W_i^{MU}, W_i^{FS} and W_i^{FNS}) will be higher. But according to Eq. (5.14) and (5.17), increasing the user effective bandwidth will decrease the user service probability when the system bandwidth is saturated, which may result in a

decrease instead in the average user throughput. Correspondingly, the average user packet delay may be higher due to lower buffer service rate in user equipment. Therefore, it is important to study the tradeoff between packet delay, time-average user bandwidth and average user throughput. Quantitative analysis on the tradeoff related to user packet delay relies on extra mathematical tools such as Queueing theory, which is beyond the scope of this work but will be explored in our future research.

5.3 Two-Level Stackelberg Game Between Macro and Femto Cells

In this section, we model the interaction between MCells and FCells into a Stackelberg game and propose a backward induction method to determine the optimal interference price y and PSU policy.

5.3.1 Game Formulation

The interaction is formulated as a two-level Stackelberg game, jointly considering the utility maximization of both MBSs and FBSs. In the first level, each MBS, as the game leader, imposes an interference-related price y upon the FBS throughput according to the interference from FBSs. In the second level, each FBS, as a follower, decides the PSU policy (i.e., n^{res} and n^{open}) based on the imposed price y, user QoS requirements and user CA capabilities. The utilities are expressed to be the total weighted profits as follows.

MBS Level Game

For each MBS, its total utility is composed of two parts: the service profits from MUs and the profits from the interference charge on FBSs. To calculate either part, it is required to have i) the average number of MUs and FBSs in one MCell (denoted as $\overline{N^{MU}}$ and $\overline{N^{FBS}}$, respectively), and ii) the average number of FSs and FNSs in one FCell (denoted as $\overline{N^{FS}}$ and

$\overline{N^{FNS}}$, respectively). Based on Eq. (5.10), $\overline{N^{MU}}$ is calculated as,

$$\overline{N^{MU}} = \sum_{k=1}^{\infty} k \cdot \mathrm{P}(N^{MU} = k). \tag{5.22}$$

Variable $\overline{N^{FBS}}$ can be calculated similarly. Based on Eq. (5.16), $\overline{N^{FS}}$ and $\overline{N^{FNS}}$ can also be achieved similarly as Eq. (5.22). Then, the utility of one MBS is given as,

$$\begin{aligned} &U^{MBS} = n_{agg} R_i^{MU} \cdot \overline{N^{MU}} g^{MU} \\ &+\omega^{MBS} y \overline{N^{FBS}} \Big[n_a R_i^{FS} \overline{N^{FS}} + n_b R_i^{FNS} \overline{N^{FNS}} \Big], \\ &where\ n_a = \min\{n_{agg}, n^{res}\},\ n_b = \min\{n_{agg}, n^{open}\} \end{aligned} \tag{5.23}$$

Here, the total throughput from all FBSs in one MCell is used to represent the interference caused by FBSs as the FBS-part interference is hard to extract from MU report in realistic implementation. ω^{MU} is the weight of interference charge over service profits. R_i^{MU} and R_i^{FS} are given in Eq. (5.6) and (5.8). As aforementioned, the MBSs can only influence the resource allocation of FBSs indirectly through price control. Therefore, one MBS can only optimize the imposed interference-related price y to maximize its own total utility:

$$\max_{0 \le y \le y^{max}} U^{MBS}. \tag{5.24}$$

FBS Level Game

For each FBS, it needs to pay g^{FS}/bit for FS services, and can gain g^{FNS}/bit for FNS services. Thus, its total utility can be expressed as:

$$\begin{aligned} &U^{FBS} = -g^{FS} n_a R_i^{FS} \overline{N^{FS}} + g^{FNS} n_b R_i^{FNS} \overline{N^{FNS}} \\ &-w^{FBS} y \Big[n_a R_i^{FS} \overline{N^{FS}} + n_b R_i^{FNS} \overline{N^{FNS}} \Big], \end{aligned} \tag{5.25}$$

where n_a and n_b are given in Eq. (5.23). Variable ω^{FBS} is the weight of interference cost over the profits. Given the interference price y imposed by the MBSs, as the user QoS re-

quirements and CA capabilities are known, the PSU policy alone can determine the effective bandwidth $W_i^T, T \in \{MU, FS, FNS\}$ and further determine the FBS utility U^{FBS}. Therefore, one FBS only needs to optimize n^{res} and n^{open} to maximize its own utility, i.e.,

$$\begin{aligned} &\max_{n^{res}, n^{open}} \quad U^{FBS} \\ &s.t. \;\; n^{res} + n^{open} \leq N, \;\; n^{res} \text{ and } n^{open} \in \mathbb{Z}^+. \end{aligned} \tag{5.26}$$

5.3.2 Analysis of the Proposed Game

Tradeoffs exist in this game. On one hand, if one MBS hopes to improve its MU performance to gain more profits from MU services, it needs to increase y to lower interference from FBSs. As a result, the throughput from FBSs will be reduced, resulting in a reduction in MBS gains from interference charge. On the other hand, one FBS can increase its utiliy by opening more carriers for FNSs, however, it needs to pay more for the increased throughput due to the interference-related price y. Therefore, MBSs need to optimize y and FBSs need to optimize the PSU policy (i.e., n^{res} and n^{open}) to achieve their own maximum utilities, i.e., to obtain the Stackelberg equilibrium.

To achieve the Stackelberg equilibrium, a backward induction method is utilized to analyze the proposed game, which captures the dependence of FBS decisions on MBS decisions. The followers of the game, i.e., the FBSs, are analyzed first. Given the imposed interference price y, the optimal PSU policy (n^{res} and n^{open}) can be achieved by solving optimization problem (5.26). The primary challenge of solving (5.26) is that the exact value of y is unknown, which means the optimal (n^{res}, n^{open}) combination is not fixed and should be a function of y. In other words, the goal of solving (5.26) is to find a mapping between different value intervals of y and the corresponding optimal (n^{res}, n^{open}) combinations. For a given y value, the general method to obtain the optimal (n^{res}, n^{open}) combination is the classic branch and bound algorithm [96], since (5.26) is typical integer nonlinear optimization. But as the backward induction method potentially needs to know the optimal (n^{res}, n^{open}) combinations for all the y values in $[0, y^{max}]$, the computation workload can be huge when N

is large. Fortunately, it is specified in the LTE-A standard [11] that at most 5 carriers can be aggregated in one system, i.e., $N \leq 5$. Therefore, there are at most 15 feasible (n^{res}, n^{open}) combinations for problem (5.26). By comparing the values of U^{FBS} under each combination within the interval $[0, y^{max}]$, the optimal (n^{res}, n^{open}) combination with the corresponding y value interval can be easily determined, as denoted below.

$$\begin{gathered} \{n_{opt}^{res}(Y_s), n_{opt}^{open}(Y_s)\}, \\ where \;\; Y_s \subset [0, y^{max}], \\ \bigcup_s Y_s = [0, y^{max}], \;\; and \;\; Y_{s_1} \bigcap Y_{s_2} = \emptyset, \\ \forall s, s_1, s_2 \in \{1, 2, \cdots, S\}. \end{gathered} \tag{5.27}$$

In Eq. (5.27), S is the total number of y value intervals that correspond to a different optimal (n^{res}, n^{open}) combination compared to its adjacent value interval.

The game for MBSs is then analyzed. As the optimal (n^{res}, n^{open} is different for different y value intervals Y_s, the utility maximization problem for MBSs (5.24) can be decomposed into a series of sub-optimization problems as follows.

$$\max_{y \in Y_s} \;\; U_s^{MBS}, \;\; s \in \{1, 2, \cdots, S\}. \tag{5.28}$$

Denote the optimal value for U_s^{MBS} and the corresponding optimal y as $U_{s,opt}^{MBS}$ and $y_{s,opt}$, respectively, then the optimal solution of the original problem (5.24) (denoted as y_{opt}) can be determined as

$$\begin{gathered} y_{opt} = y_{s^*,opt}, \\ where \;\; s^* = \arg\max_s \;\; (U_{s,opt}^{MBS}), \\ s \in \{1, 2, \cdots, S\}. \end{gathered} \tag{5.29}$$

Consequently, the optimal (n^{res}, n^{open}) combination is finalized as $\{n_{opt}^{res}(Y_{s^*}), n_{opt}^{open}(Y_{s^*})\}$.

In summary, the backward induction method obtains the Stackelberg equilibrium in two steps. It first solves the utility maximization problem of the game followers (i.e., the FBSs)

by finding a mapping between a set of y value intervals and a set of corresponding optimal (n^{res}, n^{open}) combinations. With the mapping, the utility maximization problem of game leaders (i.e., MBSs) is decomposed into a series of sub-problems with different y value intervals; by comparing the optimal utility values of each sub-problem, the optimal y for the original problem can be finalized. In this manner, the stackelberg equilibrium is determined.

5.4 Performance Evaluation

In this section, *Monte Carlo* simulation results are presented to i) validate our analytical results, and ii) demonstrate the optimal PSU policy and interference price under different user QoS requirements and CA capabilities.

5.4.1 Simulation Setup

Simulation setup of starts with an area of $20 \times 20km^2$ with $\lambda^{MBS} = 0.5/(\pi 500^2)/m^2$. The homogeneous-carrier case is considered where P_i, b_i, and α_i are identical for $\forall i$. The PSD of each MBS (FBS) is the same for every PRB in each carrier. The detailed parameter settings are presented in Table 5.1. With this setting, the average number of BSs is 255 and that of MUs is 7650. Thus, the boundary effect can be neglected by such a large-scale network. Furthermore, each presented result is averaged over 1000 runs.

5.4.2 Numerical and Simulation Results

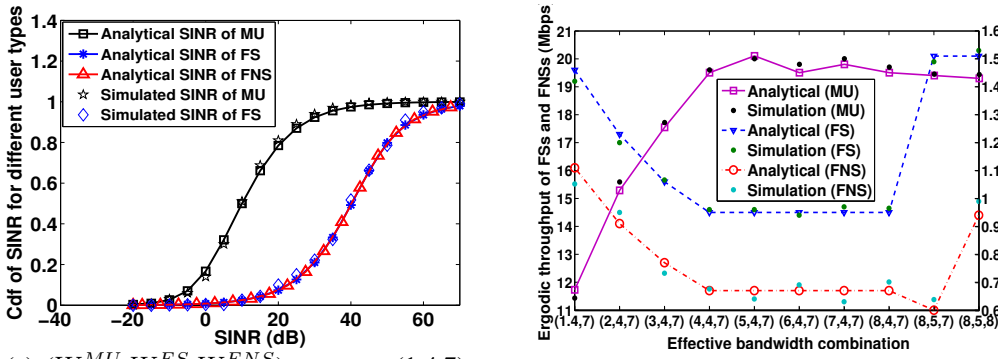
We first corroborate our analytical results on user SINR distributions and ergodic throughput in Fig. 5.3. Default values are $n_{agg} = 2$, $n^{res} = 3$, $n^{open} = 1$, $\lambda^{MU} = 30\lambda^{MBS}$, $\lambda^{FBS} = 10\lambda^{MBS}$. In Fig. 5.3(a), the cdfs of single-carrier SINR are given when the effective bandwidth of MUs, FSs and FNSs are 1,4,7, respectively. It can be observed that the SINR performance of FCell users is much better than that of MUs since MUs generally have a much longer distance to MBSs than FCell users to FBSs. Besides, FSs and FNSs have the

Table 5.1 Simulation Parameters

Parameters	Values
Coverage radius of FCell, R_F	10m
FBS density, λ^{FBS}	$10\lambda^{MBS}$
MU density, λ^{MU}	$30\lambda^{MBS}$
FS density, λ^{FS}	$3/(\pi R_F^2)/m^2$
FNS density, λ^{FNS}	$2/(\pi R_F^2)/m^2$
MBS PSD, P^{MBS}	-23.5dBm
FBS PSD, P^{FBS}	-49.5dBm
Noise PSD, n_0	$-174dBm$
Fast fading of interference	Rayleigh fading
Total number of carriers, N	5
The number of PRBs per carrier, P_i	100
PRB bandwidth in carrier i, W_{PRB}	$180kHz$
Path loss component, α_i	4
User CA capability, n_{agg}	$1 \sim 5$
MU required throughput, r^{MU}	$320 \sim 460kbps$
FS (FNS) required throughput, r^F	$5 \sim 23Mbps$
QoS violation probability, e	0.05
Unit profit of MU services, g^{MU}	10
Unit cost of FS services, g^{FS}	1.5
Unit profit of FNS services, g^{FNS}	3
(y^{max},ω^{MBS},ω^{FBS})	(10,0.01,2)

same SINR performance. This is because of the same FBS PSD for both user types and the random bandwidth access mechanism, resulting in the same strength of average useful signal and interference.

In Fig. 5.3(b), the ergodic throughput of FCell users is significantly higher ($\sim$ 10 times) than that of MUs under different effective bandwidth combinations due to much better SINR performance. It can be further observed that when each MU is assigned with more PRBs per carrier, the MU ergodic throughput first increases and then remains stable. This can be explained as follows: when W_i^{MU} is small, increasing W_i^{MU} will bring each MU more bandwidth without increasing the interference intensity from other MBSs very much. Thus, the ergodic throughput increases. However, if W_i^{MU} keeps increasing, the service probability of each MU ($Q_{s|i}^{MU}$) will drop considerably, which counterbalances the performance gain brought by wider bandwidth. So the ergodic throughput becomes stable. In addition, FSs have a higher throughput than FNSs since FSs can concurrently transmit on 2 carriers ($n_{agg} = 2$ and $n^{res} = 2$) compared to 1 carrier for FNSs. When W_i^{MU} increases, the ergodic



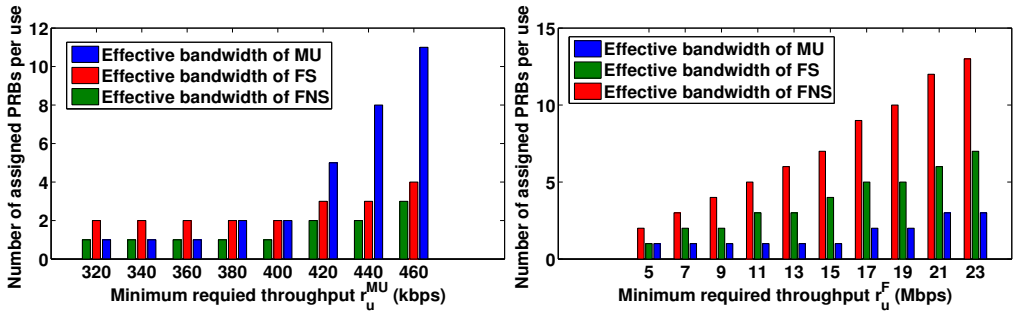
(a) (W_i^{MU},W_i^{FS},W_i^{FNS}) are set to (1,4,7) PRBs.

(b) Total ergodic throughput per user.

Figure 5.3. User SINR and ergodic throughput performance in HetNets.

throughput of FSs and FNSs both decreases first and then becomes stable. This is because with larger W_i^{MU}, the interference from MBSs first increases and then remains unchanged since the bandwidth usage probability $\theta_i^{MBS,usa}$ has reached its maximum, i.e., 1. Moreover, when W_i^{FS} increases, e.g., from (8,4,7) to (8,5,7), the throughput of FNS decreases, which is because the interference perceived by FNSs from FBSs increases.

Fig. 5.4 shows how the effective bandwidth of different user types is decided with the user throughput requirements given QoS violation probability e and CA capabilities. Default values are the same with Fig. 5.3. The effective bandwidth is represented by the number of PRBs assigned to each user. It can be seen that as the throughput requirements increase, users need to be assigned with more PRBs to satisfy the maximum QoS violation probability. Besides, MUs are more sensitive to the throughput increase than FSs and FNSs: W_i^{MU} increases 10 times to satisfy only 44% increase of r^{MU} while W_i^{FS} or W_i^{FNS} increases 7 times to satisfy 360% increase of r^F. The reason is as follows. In an MBS, there are more MUs selecting the same carrier than FSs (or FNSs) do in an FBS, resulting in that the bandwidth in one carrier is more likely to be saturated in an MBS than in an FBS. Thus, increasing W_i^{MU} will more likely reduce the user service probability of MUs, making the MU throughput increase harder than FSs and FNSs. Furthermore, it can be observed in Fig. 5.4(a) that even if r^F is not changed, the effective bandwidth of FSs and FNSs still increases to supplement the throughput loss due to increased interference from MBSs. The similar



(a) r_u^{MU} changes while $r_u^F = 5Mbps$. (b) r_u^F changes while $r_u^{MU} = 360kbps$.

Figure 5.4. Effective bandwidth with different minimum throughput requirements.

phenomenon is also observed in Fig. 5.4(b).

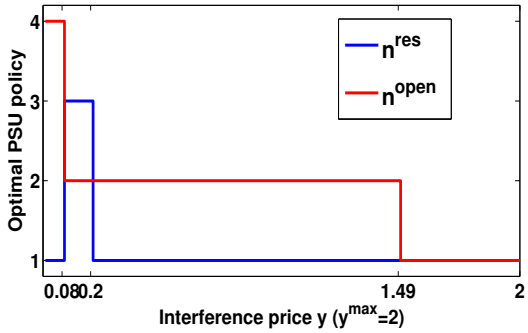


Figure 5.5. The optimal PSU policy of FBS (n^{res} and n^{open}) given different y.

Finally, we show how the optimal PSU policy and interference price y is determined with given user throughput requirements and CA capabilities. As mentioned in Subsection 5.3.2, to maximize the utilities of both parties, the FBSs first offer MBSs with the knowledge of optimal PSU policies to maximize the utility of FBS given different y values. Then the MBS choose the optimal y to maximize its own utility. Fig. 5.5 shows the mapping between optimal PSU policy and y when $n_{agg} = 4$, $r^{MU} = 400kbps$, and $r^F = 15Mbps$. It can be seen that the optimal PSU policy changes when y reaches the point 0.08, 0.2 and 1.49. As y increases, the total number of carriers selected by an FBS decreases in order to reduce the interference cost charged by MBSs.

The optimal interference price and PSU policy for different CA capabilities is evaluated in Fig. 5.6 with $r^{MU} = 400kbps$ and $r^F = 15Mbps$. When $n_{agg} = 3, 4$, the optimal

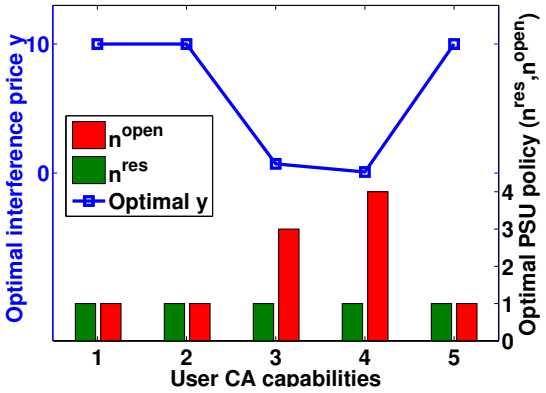


Figure 5.6. The optimal interference price and PSU policy for different CA capabilities.

interference price is 0.71 and 0.08, respectively. This implies that for FBSs, the increase of utility due to enlarging the number of carriers for FNS surpasses the resultant interference cost charged by MBSs. When $n = 1, 2$ and 5, the situation is adverse where the FBSs have to reduce the number of open-access carriers to the minimum to avoid the relatively excessive interference charges from MBSs.

5.5 Summary

In this chapter, we have studied the QoS provisioning in LTE-A HetNets with PSU mechanism. Specifically, the Stochastic geometry has been leveraged to consider the random behaviors of the FCells and inter-macro interference into performance analysis under PSU mechanism. Then, the concept of effective bandwidth has been applied to decide the user bandwidth with considering the user QoS requirements and CA capabilities. Furthermore, the interaction between MBSs and FBSs has been modelled into a Stackelberg game to maximize the utilities of both parties. Finally, *Monte Carlo* simulations have been performed to verify our analytical results and demonstrate the decision process of effective bandwidth as well as the optimal PSU policy and interference price. The research outcomes should shed some light on how to optimally coordinate the resource utilization in HetNets among different operator bands, which is a future trend for the cellular systems.

Chapter 6

Modeling and Analysis of MAC Protocol for LTE-U Coexisting with Wi-Fi

In this chapter, we propose a MAC protocol for LTE-U and analyze the performance of the co-existed LTE-U and WiFi networks. The main contributions are three folds. First, we propose an Listen-before-Talk (LBT) based LTE-U MAC protocol to achieve harmonious co-existence with Wi-Fi systems. Specifically, the MAC timing is divided into variable cycles, each composed of variable periods for LTE-U transmissions, WiFi transmissions, and channel sensing. By adjusting the ratio between the transmission periods of LTE-U and Wi-Fi adaptively, performance balance between the LTE-U and Wi-Fi can be ensured and the desired Wi-Fi protection levels can be achieved. Second, based on the proposed MAC, we develop an analytical model to study the network performance of LTE-U and Wi-Fi, capturing the asynchronous nature of Wi-Fi in a synchronous MAC frame structure of LTE-U. We analyze the average duration of channel sensing in each cycle for LTE-U to retrieve the channel access, and derive the average throughput of both LTE-U and Wi-Fi networks. Finally, extensive simulations are conducted to validate the throughput analysis under different sensing configurations and Wi-Fi protection levels, which provides important guidance for the optimal MAC setting in an LTE-U/Wi-Fi co-existing system.

6.1 System Model

6.1.1 Co-existence Scenario

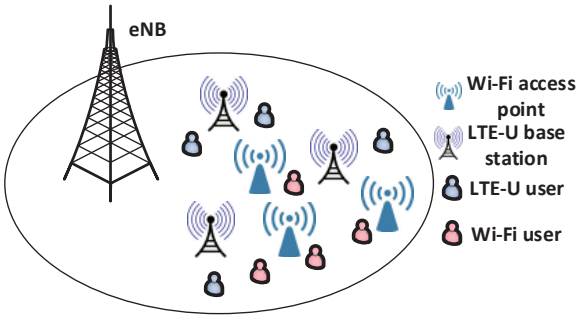


Figure 6.1. Coexistence scenario between LTE-U and Wi-Fi.

Fig. 6.1 shows a network with co-existed LTE-U and Wi-Fi operating on the same unlicensed spectrum band. Due to the unlicensed transmission power limitations [73], the LTE-U technology will be mainly used for small cells, yet the small cells may operate on both licensed and unlicensed bands. Data with high reliability and QoS requirements, e.g., control signaling, is transmitted over the licensed bands; while other supplemental data can be transmitted over the unlicensed bands [75]. In a dense deployment, it is possible that more than two different access networks, e.g., LTE-U small cells or Wi-Fi networks, select the same channel, and cause inter-system interference. Unlike the traditional LTE system where macro- and micro-cells are managed by one operator for efficient coordination, coordination is difficult for unlicensed networks of different operators or Wi-Fi owners. Therefore, it is critical to design a co-existence mechanism to allow efficient spectrum sharing between LTE-U and Wi-Fi.

6.1.2 LTE and Wi-Fi MAC/PHY Features

The LTE and Wi-Fi have different PHY and MAC features. LTE employs orthogonal frequency division multiple access (OFDMA) in the PHY layer. The whole bandwidth is divided into a set of orthogonal physical resource blocks (PRBs). Different PRBs can be sched-

uled to different users in the same subframe, thus achieving multi-user diversity gain. Wi-Fi adopts orthogonal frequency division multiplexing (OFDM) in the PHY layer, but only one user can access the channel at one time. For the MAC protocols, LTE adopts a centralized and synchronous MAC to schedule transmissions in each subframe of 1 ms. Wi-Fi uses a distributed asynchronous MAC based on CSMA/CA [97]. That is, before transmission, the Wi-Fi node first listens to the intended channel. If the channel is busy, the Wi-Fi node will backoff for a random time to reduce collision probability. Due to these differences, MAC design for LTE-U should be based on the synchronous structure of LTE for easy integration and compatibility, yet should also be efficient and adaptive to the asynchronous Wi-Fi transmissions.

6.2 The Proposed LBT-Based MAC for LTE-Unlicensed

The basic principle of the LBT-based MAC is that LTE-U nodes need to sense the channel for a period before transmission. If the channel is sensed busy, the LTE-U node should remain silent and sense the channel periodically in the following subframes till the channel is idle for a certain duration. In 3GPP meetings, it is generally accepted that alternating channel reservation periods for LTE-U and Wi-Fi should be adopted in the LTE-U MAC [98]. Based on these principles [98], we propose a detailed LTE-U MAC protocol to coordinate the unlicensed spectrum sharing with Wi-Fi systems, as illustrated in Fig. 6.2. Parameters of the MAC protocol play a critical role in the performance of coexisted networks, and will be analytically studied in Section 6.3.

Timing in LTE-U is slotted into subframes of 1ms, as shown in Fig. 6.2. Several subframes are reserved for LTE-U and Wi-Fi transmissions, respectively, i.e., the LTE-U transmission period (LTX, the blue period) and the Wi-Fi transmission period (WTX, the red region). Both LTX and WTX can be adjusted according to the desired performance of either system. In the last subframe of WTX, the LTE-U node can start channel sensing. The subframes where sensing is performed are called sensing subframes (SSs), and the duration

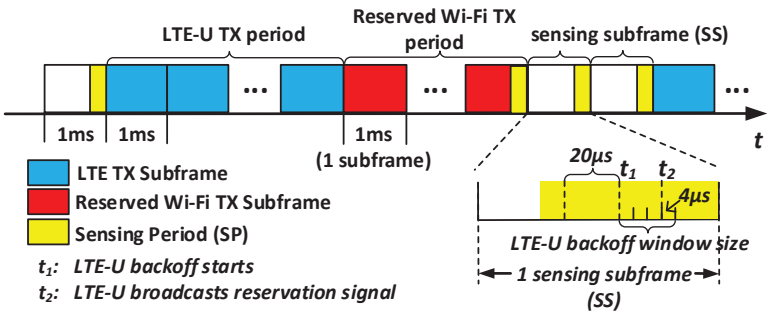


Figure 6.2. LBT-based MAC protocol of LTE-U.

of sensing in each SS is called one sensing period (SP) as marked in yellow at the end of each SS. If the channel is sensed idle for a duration, the LTE-U node broadcasts a reservation signal to reserve the next few subframes for LTX; otherwise, the node will sense in every following SS until the channel is sensed idle. The LTE-U node can only transmit at the beginning of the subframe following the reservation signals, while Wi-Fi can transmit at any time during a subframe. Due to the asynchronous Wi-Fi transmissions, there may be a variable number of SSs before LTX starts. The duration of the SPs affects the network throughput of both LTE-U and Wi-Fi.

The sensing procedure is illustrated in Fig. 6.2. The SP has a minimum duration of $20\mu s$. In one SP, if the channel is sensed idle for $T_{thre}\mu s$, the LTE-U node adds a random backoff time to avoid reservation collisions with different LTE-U nodes. The backoff timer elapses when the channel is sensed idle. When the backoff timer reaches 0, the LTE-U node immediately broadcasts a reservation signal. If a new Wi-Fi transmission or a reservation signal from another LTE-U occurs during the backoff, the reservation fails and the above process has to be repeated till a successful reservation is launched.

It is also interesting to point out that LBT sensing may make it difficult for LTE-U to retrieve the channel access from Wi-Fi. Unlike the Wi-Fi nodes that can sense the channel continuously, the channel sensing of LTE-U is usually performed at specific time in one SS. When the Wi-Fi traffic load is medium or high, it is with high probability that there are ongoing Wi-Fi transmissions covering the SP due to the elastic and asynchronous channel access

nature. Consequently, the channel will be sensed busy with high probability, resulting failure of the LTE-U system to retrieve the channel access. In such case, LTE-U may achieve a low throughput while Wi-Fi is well protected. Therefore, the duration of SP can significantly affect the LTE-U success probability for channel retrieval, and further affect the throughput performance of both systems.

In the proposed LTE-U MAC, we set $T_{thre} = 20\mu s$, and the maximum LTE-U backoff timer is 3 slots with a slot duration of $4\mu s$. Notice that a DCF interframe space (DIFS) duration is $34\mu s$, i.e., a Wi-Fi transmission can be initiated after the channel is sensed idle for at least $34\mu s$. By allowing an LTE-U node to reserve LTX transmissions after a maximum of $20\mu s + 3 * 4\mu s = 32\mu s$, LTE-U node has a higher priority for channel access compared with Wi-Fi nodes during the SP so that LTE-U can easily retrieve the channel access after the WTX completes. This can ensure the throughput efficiency of LTE-U while providing the satisfactory Wi-Fi performance via WTX reservations. The main notations of the proposed protocol are summarized in Table 6.1.

6.3 Performance Analysis for LTE-U LBT-Based Coexistence Mechanism

This section presents throughput analysis for the network with one LTE-U small cell co-existed with Wi-Fi networks. Firstly, the average number of SPs taken by LTE-U to retrieve the channel access is derived. Then, the average system throughput is derived for both LTE-U and Wi-Fi systems. Finally, discussions are provided on how to tune the MAC parameters according to the desired Wi-Fi protection level.

Table 6.1 Notation Table

Notations	Physical Meanings
T_s	The duration of one LTE subframe, $1ms$*
LTX	The reserved period for LTE transmissions in one cycle
WTX	The reserved period for Wi-Fi transmissions in one cycle
$\overline{WTX_{tot}}$	Average total period that the Wi-Fi system can transmit consecutively
T_{SP}	Channel sensing period in one sensing subframe (SS)**
$\overline{T_{int}}$	The time interval between two consecutive Wi-Fi transmissions
T_{DIFS}	The duration of DIFS in Wi-Fi MAC protocol*
$T_{W,Win}$	The maximum contention window size of Wi-Fi**
$T_{L,Win}$	The maximum backoff window size of one LTE-U node**
T_{thre}	The minimum sensing time threshold to determine whether channel is idle, i.e., $20\mu s$*
$APWT$	The average period per Wi-Fi transmission**
$\overline{R_L}$	The average cell throughput of one LTE-U system
$\overline{R_W}$	The average sum throughput of the Wi-Fi system
$\overline{N_{SP}}$	The average number of SPs needed for LTE-U to get the channel access back
Remarks	The variables marked with * are constants. The variables marked with ** are configurable parameters.

6.3.1 Average Number of Sensing Periods to Retrieve the Channel Access

Denote the average number of SPs needed by the LTE-U to retrieve the channel access as $\overline{N_{SP}}$. $\overline{N_{SP}}$ is closely dependent on the position of the first Wi-Fi transmission that overlaps with the first SS of LTE-U. Such Wi-Fi transmissions are referred to as *first overlapping (FO) transmissions* in this work. There are two types of FO transmissions as illustrated in Fig.6.3 : i) those that start before the first SS and stretch into the SSs, and ii) those that start in the first SS but their respective previous Wi-Fi transmission ends before the first SS. As specified in 802.11 standard [99], when one Wi-Fi transmission ends, other nodes first sense the channel for a DIFS duration (denoted as T_{DIFS}) and transmit only when the backoff timer reaches 0. Thus the average interval between two consecutive Wi-Fi transmissions, denoted as $\overline{T_{int}}$, is

$$\overline{T_{int}} = T_{DIFS} + T_{W,Win}/2, \tag{6.1}$$

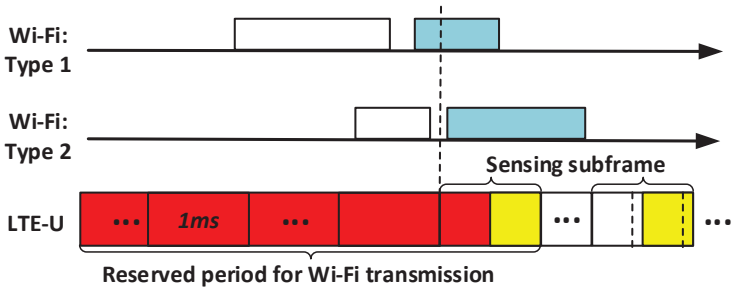


Figure 6.3. Illustration of two types of FO Wi-Fi transmissions.

where $T_{W,Win}$ is the maximum random backoff window size in μs. Thus, the average range where the second-type FO transmissions start is $\overline{T_{int}}$. Therefore, the total range where FO transmissions start is

$$T_{range} = APWT + \overline{T_{int}}, \tag{6.2}$$

where $APWT$ is the average period per Wi-Fi transmission. Meanwhile, to retrieve the channel access, one LTE-U node needs to first sense channel for a T_{thre} period plus back-off time in SP before broadcasting the reservation signal. Thus, the average period from the instance when the channel is sensed idle to the instance when the reservation signal is broadcast is

$$T' = T_{thre} + T_{L,Win}/2. \tag{6.3}$$

where $T_{L,Win}$ denotes the maximum backoff window size for one LTE-U node in μs.

For FO transmissions, different positions where the transmission completes can lead to different channel sensing results and different numbers of following SSs. Thus, we study the four cases with different positions of the FO transmissions, as illustrated in Fig. 6.4. To simplify the analysis, we consider that APWT is a multiple of milliseconds.

Case 1: If the Wi-Fi transmission ends no earlier than $(\overline{T_{int}} - T')$ μs before SP begins and no later than T_{thre} μs before SP ends, the LTE-U node, e.g., LTE-U base station (BS), is able to detect a T_{thre} idle period and broadcast the reservation signal before another Wi-Fi transmission starts. In this case, no more SPs are needed and LTE can start transmission in the beginning of the next subframe.

Case 2: If the Wi-Fi transmission ends more than $(\overline{T_{int}} - T')\mu s$ before the SP starts, the LTE-U node cannot detect a T_{thre} μs idle period in the current SS or has no enough time to broadcast the reservation signal. In this case, a new Wi-Fi transmission may start and the LTE-U node has to sense at least another $APWT/T_s$ SSs to retrieve the channel access.

Case 3: If the Wi-Fi transmission ends later than T_{thre} μs before the SP ends, the LTE-U node cannot detect a T_{thre} μs idle period in the current SS. In this case, the LTE-U node has to sense at least another $APWT/T_s + 1$ subframes to get the channel access back.

Case 4: Case 4 is different from the above ones in the sense that it corresponds to the second type of FO transmissions, i.e., the FO transmissions that start in the first sensing subframe but their respective previous transmission ends before the first sensing subframe. In this case, the LTE-U node cannot detect any idle period in the first SS in an average sense and has to sense at least $APWT/T_s + 1$ subframes before transmitting.

In the following, we will calculate the average number of SPs before an LTX can be launched, i.e., $\overline{N_{SP}}$, for different cases. Denote $\overline{N_{SP}}$ for case c as $\overline{N_{SP_c}}, c \in \{1, 2, 3, 4\}$. We assume that the arrival of the FO transmissions follows a uniform distribution within T_{range}.

Case 1

In a single subframe, the probability that one FO transmission belongs to case 1 is calculated as

$$p_1 = [(T_{SP} - T_{thre}) + (\overline{T_{int}} - T')]/T_{range}. \tag{6.4}$$

If one FO transmission corresponding to case 1 starts within the ith subframe of the first APWT duration in T_{range}, as one transmission lasts for APWT duration on average, the average number of SPs needed to find the transmission end is i. Thus $\overline{N_{SP_1}}$ is calculated as

$$\begin{aligned} \overline{N_{SP_1}} &= \sum_{i=1}^{APWT/T_s} p_1 \cdot i \\ &= \frac{\overline{T_{int}} - T' + T_{SP} - T_{thre}}{T_{range}} \cdot \frac{(1 + \frac{APWT}{T_s})APWT}{2T_s} \end{aligned} \tag{6.5}$$

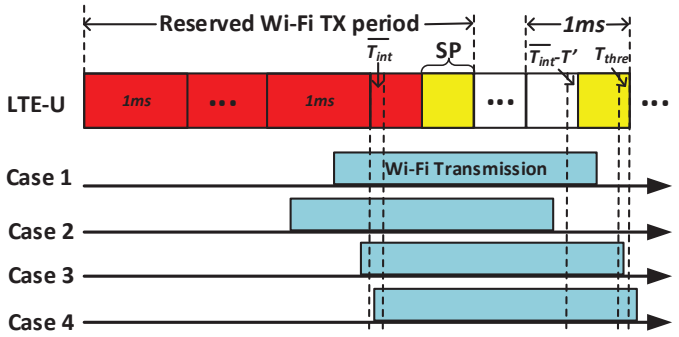


Figure 6.4. Illustration of four cases for FO transmissions.

Case 2

In one subframe, the probability that one FO transmission belongs to case 2 is given by

$$p_2 = [(T_s - T_{SP}) - (\overline{T_{int}} - T')]/T_{range}. \tag{6.6}$$

In this case, the LTE-U node cannot detect an idle period of T_{thre} μs or cannot finish the backoff to broadcast a reservation signal. Only Wi-Fi transmissions may start and last for APWT on average until a new Wi-Fi transmission completes in the region corresponding to case 1. Thus according to the number of new transmissions that will occur, the region corresponding to case 2 in one SS is divided into N_{sr} sub-regions, where

$$N_{sr} = \left\lceil \frac{T_{C_2}}{\overline{T_{int}}} \right\rceil, T_{C_2} = (T_s - T_{SP}) - (\overline{T_{int}} - T_{thre}). \tag{6.7}$$

The conditional probability that the transmission ends in the the first sub-region given that it belongs to case 2 is

$$p(\text{first sub-region}|C_2) = (T_{C_2} - N_{sr} \cdot \overline{T_{int}})/T_{C_2}. \tag{6.8}$$

The conditional probability that the transmission ends in one of the other sub-regions given that it belongs to case 2 is

$$p(\text{another sub-region}|C_2) = \overline{T_{int}}/T_{C_2}. \tag{6.9}$$

Eq. (6.8) and (6.9) are different because the first sub-region may be less than $\overline{T_{int}}$. For the jth sub-region, the LTE-U node needs to sense $(N_{sr}+1-j)APWT/T_s$ more SPs before transmitting. Thus, the average number of extra SPs due to new Wi-Fi transmissions, denoted as $\overline{N_{SP_2}^{new}}$, is calculated as

$$\begin{aligned}\overline{N_{SP_2}^{new}} = \tfrac{APWT}{T_s}[&p(\text{first sub-region}|C_2)N_{sr} \\ &+ \sum_{j=2}^{N_{sr}} p(\text{another sub-region}|C_2)(N_{sr}+1-j)].\end{aligned} \tag{6.10}$$

Then, the average total number of SPs needed due to both current and new Wi-Fi transmissions is calculated as

$$\overline{N_{SP_2}} = \sum_{i=1}^{APWT/T_s} p_2 \cdot (i + \overline{N_{SP_2}^{new}}). \tag{6.11}$$

Case 3

The way of calculation for case 3 is similar to that of case 2, i.e., exploiting sub-regions to calculate the average number of extra SPs due to new Wi-Fi transmissions. We divide the combined region, i.e., the region corresponding to case 2 in the current SS and the region corresponding to case 3 in the next SS, into N'_{sr} sub-regions, i.e.,

$$N'_{sr} = \left\lceil \frac{T_{C_2}+T_{thre}}{\overline{T_{int}}} \right\rceil, \tag{6.12}$$

where T_{R_2} is given in Eq. (6.7). The region corresponding to case 3 may belong to the same sub-region or two sub-regions. If $(N'_{sr}-1)\overline{T_{int}} \leq T_{C_2}$, the entire region of case 3 in one single subframe belongs to the 1st sub-region, otherwise the region stretches across the 1st and 2nd sub-regions. Therefore, the average number of SPs due to new Wi-Fi transmissions,

denoted as $\overline{N_{SP_3}^{new}}$, is calculated as

$$\begin{aligned}&\overline{N_{SP_3}^{new}} = 1 + APWT/T_s \\ &\cdot \begin{cases} N'_{sr}, & \text{if } (N'_{sr}-1)\overline{T_{int}} \leq T_{C_2}; \\ \frac{T_{C_2}+T_{thre}-(N'_{sr}-1)\overline{T_{int}}}{T_{thre}} N'_{sr} & \\ \quad + \frac{(N'_{sr}-1)\overline{T_{int}}-T_{C_2}}{T_{thre}}(N'_{sr}-1), & \text{otherwise.} \end{cases}\end{aligned} \tag{6.13}$$

Thus the average total number of SPs due to both current Wi-Fi transmissions and new Wi-Fi transmissions is

$$\overline{N_{SP_3}} = \sum_{i=1}^{APWT/T_s} \frac{T_{thre}}{T_{range}} \cdot (i + \overline{N_{SP_3}^{new}}). \tag{6.14}$$

Case 4

Similarly, we can divide the sub-regions in case 4. Divide T_{C_2} by $\overline{T_{int}}$ and the duration of the region for case 4 is $\overline{T_{int}}$. The region for case 4 must stretch across the 1st and 2nd sub-regions. Thus the average number of SPs needed in case 4, i.e., $\overline{N_{SP_4}}$, is calculated as

$$\begin{aligned}&\overline{N_{SP_4}} = 1 + \frac{APWT \cdot \overline{T_{int}}}{T_{range}} \cdot \\ &\left[\frac{T_{C_2}-(N_{sr}-1)\overline{T_{int}}}{\overline{T_{int}}}(N_{sr}+1) + \frac{N_{sr}\overline{T_{int}}-T_{C_2}}{\overline{T_{int}}} N_{sr}\right].\end{aligned} \tag{6.15}$$

Therefore, combining Eq. (6.5)(6.11)(6.14)(6.15), the average number of SPs needed before a LTX is

$$\overline{N_{SP}} = \sum_{c=1}^{4} \overline{N_{SP_c}}. \tag{6.16}$$

6.3.2 Average System Throughput for LTE-U and Wi-Fi Systems

After obtaining the average number of SPs needed for an LTE-U node to retrieve the channel access, i.e., $\overline{N_{SP}}$, we then derive the network throughput of the coexisted LTE-U and Wi-Fi. During the WTX and the SSs, LTE-U cannot launch a transmission. Denote the total average

duration of this period that LTE-U cannot transmit as $\overline{T_{NL}}$,

$$\overline{T_{NL}} = WTX - T_s + \overline{N_{SP}} \cdot T_s. \tag{6.17}$$

One "T_s" in Eq. (6.17) was reduced because the first SP starts in the last subframe of the reserved Wi-Fi transmission period WTX. Note that $\overline{T_{NL}}$ does not equal to $\overline{WTX_{total}}$ which denotes the average duration that the Wi-Fi system can transmit consecutively during each LTE-U transmission cycle. This is because the Wi-Fi cannot transmit after the LTE-U node broadcasts the reservation signal in the last SS. Therefore, $\overline{WTX_{total}}$ should be calculated as

$$\begin{aligned}&\overline{WTX_{total}} = \\ &\overline{T_{NL}} - T_s + p_1 \underbrace{\frac{T_{C_2} + (T_s - T_{thre})}{2}}_{A_1} + (1 - p_1) \underbrace{T_{C_2}}_{A_2}.\end{aligned} \tag{6.18}$$

In Eq. (6.18), $p_1 A_1 + (1 - p_1) A_2$ represents the average duration that Wi-Fi transmits in the last SS due to the asynchronous transmission nature of Wi-Fi. A_1 denotes the average duration that one Wi-Fi transmission belonging to case 1 can last. The part A_2 corresponds to case 2, 3, and 4, because the transmissions belonging to cases 2, 3, and 4 end very close to the beginning of the SP in the last SS. Ignoring the time differences in the order of μs, we use a unified duration T_{C_2} for these three cases.

Given LTX, $\overline{T_{NL}}$ and $\overline{WTX_{total}}$, the system throughput for LTE-U and Wi-Fi can be derived. For the LTE-U system, only the throughput on unlicensed band is presented since our concern is the coexisting performance on unlicensed spectrum. Following [35], the average throughput of an LTE cell when proportional-fair MAC scheduling and adaptive power allocation are adopted is given by

$$\begin{aligned}&\overline{R_{LTE}} = \frac{N_{LTE} N_{PRB}}{\sum_{i=1}^{N_{LTE}} R_i^{-1}}, \\ &\text{where } R_i = W_{PRB} \log_2 \left[1 + \frac{\Gamma \cdot 10^{(\alpha-1)L_i/10}}{N_0}\right], \\ &\qquad \Gamma = 10^{(P_0 - 30)/10},\end{aligned} \tag{6.19}$$

where N_{LTE} denotes the number of LTE users, N_{PRB} the number of PRBs in one channel, W_{PRB} the bandwidth of one PRB, L_i the pass loss of user i, N_0 the noise power per PRB, and (α,P_0) the frequency-specified power control parameters. Due to time sharing, the actual average throughput of LTE-U in coexisted scenario is

$$\overline{R_L} = \overline{R_{LTE}} \frac{LTX}{LTX + \overline{T_{NL}}}. \tag{6.20}$$

For a Wi-Fi system, as given in [100], the aggregate network throughput can be derived as

$$\begin{aligned} &\overline{R_W} = \beta \cdot \frac{P_{suc} S_{pay}}{P_{idle} T_{W,Win}/2 + P_{suc} APWT + P_{col} T_{col}}, \\ &\text{and } \beta = \frac{\overline{WTX_{total}}}{LTX + \overline{T_{NL}}}, \end{aligned} \tag{6.21}$$

where S_{Pay} denotes the average number of bits per Wi-Fi transmission, T_{col} denotes the average duration of a collision period, P_{suc}, P_{idle} and P_{col} are the probabilities of a successful transmission, an idle slot, and a collision, which are determined by the analytical framework in [100].

From Eq. (6.20) and (6.21), it can be seen that the level of Wi-Fi performance protection is closely related to the ratio β, which is dependent of WTX and T_{SP}. Thus, the following guidelines are provided in Proposition 1 to show how to tune WTX based on LTX, T_{SP} and the required Wi-Fi performance protection level.

Proposition 1 *Given the T_{SP} and LTX, to protect η percent of Wi-Fi system throughput, WTX should be set that the following condition is satisfied,*

$$\frac{\overline{WTX_{total}}}{LTX + \overline{T_{NL}}} \geq \eta\% \tag{6.22}$$

As the corresponding equality of Eq. (6.22) is a transcendental equation without an analytical solution, numerical results is presented in next subsection to show how WTX changes with different η and T_{SP} values.

6.4 Performance Evaluation

In this section, simulation results are provided to validate our analysis and demonstrate the performance of the proposed LTE-U MAC.

We simulate the coexistence network scenario presented in Fig. 6.1, where one LTE-U small cell coexists with a set of Wi-Fi access points (APs) on a $20MHz$ unlicensed channel. The simulation runtime is 100s. For the LTE-U system, two LTE-U users are uniformly located within the cell with a radius of $40m$ with only downlink transmissions. For the Wi-Fi system, all the APs and users are uniformly distributed within the LTE-U cell. All the nodes carry saturated traffic. The main simulation parameters are shown in Table 6.2.

Table 6.2 Main Simulation Parameters

LTE-U Parameters	**Values**
Path Loss Factor	3
Shadowing Statistics	$\mu = 0\ dB$, $\sigma = 8\ dB$
Noise Power per RB, N_0	$-90dBm$
$[\alpha,\ P_0]$	$[0.5, -58dBm]$
Power allocation for user i (dBm)	$P_0 + 10\log(N_{PRB}) + \alpha L_i$
Number of PRBs in the channel	100
Bandwidth per PRB	$180kHz$
MAC scheduling method	Proportional Fair
LTX	$5ms$
Default sensing period T_{SP}	$200\mu s$
Channel idle time threshold T_{thre}	$20\mu s$
Maximum Backoff window size	3 slots
Backoff time slot duration	$4\mu s$
Wi-Fi Parameters	**Values**
Default WTX	$5ms$
DIFS duration T_{DIFS}	$34\mu s$
Maximum contention window size	31 slots
Backoff time slot duration	$9\mu s$
Packet size S_{pay}	6750 bytes

We first investigate the probability that the LTE-U BS fails to reserve the channel in the first SS under different T_{SP} in Fig. 6.5. WTX is set to 5ms. It can be seen that when T_{SP} is very small, i.e., $20\mu s$, the failure probability can be as high as 0.82. This is because unlike the Wi-Fi system that can sense the channel continuously, the LTE-U BS senses the channel at specific time, i.e., at the last $20\mu s$ in one SS. As a result, the probability that the BS

can detect a $20\mu s$ idle period and broadcast the reservation signal is very small for a lower T_{SP}. When T_{SP} increases, the failure probability is reduced since the BS is more likely to detect a $20\mu s$ idle period and launch a successful reservation with a larger SP duration. The

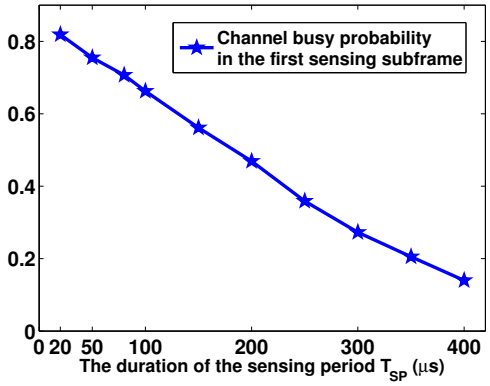


Figure 6.5. Failure probability for the LTE-U BS to reserve the channel in one subframe.

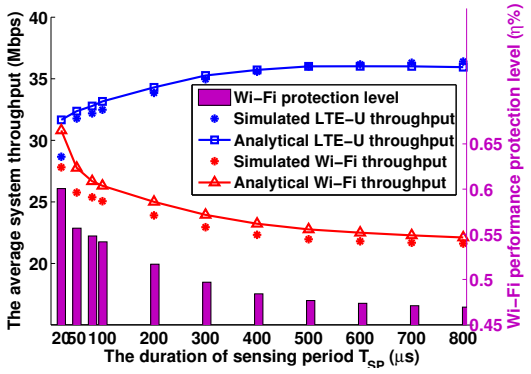


Figure 6.6. Average system throughput of LTE-U and Wi-Fi with Wi-Fi protection level.

analytical and simulated results of average system throughput are shown in Fig. 6.6 for both systems with WTX=5ms. It can be observed that when T_{SP} increases, the average Wi-Fi throughput decreases while that of the LTE-U system increases. Meanwhile, the protection level of the Wi-Fi performance reduces when T_{SP} increases. The reason is as follows. When T_{SP} increases, the probability that the LTE-U BS can successfully reserve the channel in one SS becomes larger. Consequently, the average number of SSs for the LTE-U BS to retrieve the channel is smaller. As Wi-Fi can still transmit in the SS, the reduced number of SSs will result in reduced time ratio of the Wi-Fi transmission in the whole simulation period. Thus, the Wi-Fi throughput decreases and the LTE-U throughput increases. Besides, the analytical results match well with the simulation results. The gap between the analysis and simulation is slightly larger when T_{SP} is smaller. This is because the analysis assumes that the end of the Wi-Fi transmissions belonging to case 1 follows a uniform distribution. When T_{SP} is smaller, the assumption becomes less accurate, resulting a larger gap.

Finally, we show that given LTX and T_{SP}, how to adjust WTX according to different Wi-Fi protection levels in Fig. 6.7. For a given T_{SP}, the average number of SSs needed by

LTE-U to retrieve the channel can be determined. Thus, to increase the Wi-Fi protection level is equivalent to increase the time ratio of Wi-Fi transmissions in one cycle. As shown in the figure, the average Wi-Fi throughput increases with a larger η at the expense of the decreased average LTE-U throughput because a smaller ratio of time is allocated for LTE-U transmissions.

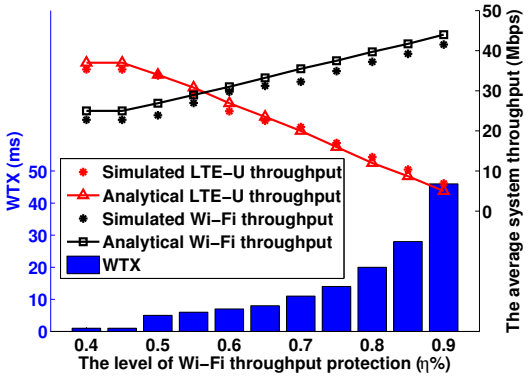


Figure 6.7. Adjustment of WTX to different Wi-Fi protection levels with $T_{SP} = 200\mu s$.

6.5 Summary

In this chapter, we have presented a new MAC protocol design for LTE-U networks. An analytical model has been developed to study the MAC throughput performance of co-existed LTE-U and Wi-Fi networks, considering the asynchronous WiFi transmissions in a time slotted MAC structure of LTE-U. It has been shown that a certain level of Wi-Fi protection can be achieved by adaptively adjusting the MAC parameters. In our future work, we will extend our performance study by considering multiple co-existed small cells and Wi-Fi networks operating in multiple unlicensed channels.

Chapter 7

Conclusions and Future Works

7.1 Conclusions

In this dissertation, we have introduced the LTE-A systems and the features of CA. The research topics and related works of CA-aware RRM in LTE-A systems are fully elaborated. In addition, our research works are presented where enhanced CA-aware RRM strategies are developed for single-tier LTE-A systems, multi-tier LTE-A systems and LTE-U systems with licensed and unlicensed spectrum aggregation.

Firstly, we have studied the EC performance of LTE-A systems with CA in downlink admission control process to demonstrate the benefits of CA. Closed-form expressions of EC have been derived with the binomial-normal approximation for both kinds of users under two different bandwidth allocation strategies. We have further formulated a net-profit-maximization problem to investigate the tradeoff among the bandwidth weights for heterogeneous user classes. Finally, extensive simulations have demonstrated the accuracy of our analytical results and shown that with only a small increase in the spectrum utilization, LTE-A users can have considerably higher EC than LTE users when the user traffic is bursty.

Secondly, we have investigated the cross-layer RRM performance of uplink CA in LTE-A systems. A joint CC selection and power control strategy is proposed to enhance the average user throughput, considering the user power constraints and offset effects. The results sug-

gest that assigning more CCs to a user is not always preferable since multi-CC transmission will increase the interference intensity on each CC and cause additional non-transmission power consumption for user device. Thus, the CC selection should be closely dependent on the cell load conditions.

Thirdly, we have explored the QoS provisioning in LTE-A HetNets with PSU mechanism. In the work, we exploit the Stochastic geometry to depict the random behaviors of the FCells and inter-macro interference under PSU mechanism. We then model the interaction between MBSs and FBSs into a Stackelberg game to maximize the utilities of both parties. The research outcomes should shed some light on how to optimally coordinate the resource utilization in HetNets among different operator bands, which is a future trend for the cellular systems.

Last, we have presented a new MAC protocol design for LTE-U networks. An analytical model has been developed to study the MAC throughput performance of co-existed LTE-U and Wi-Fi networks, considering the asynchronous WiFi transmissions in a time slotted MAC structure of LTE-U. By adaptively adjusting the MAC parameters, a certain level of Wi-Fi protection can be achieved. Our design jointly considers the instantaneous channel occupation conditions (i.e., on-going Wi-Fi transmissions) and the channel retrieving difficulty for the LTE-U BSs, through which the harmony coexistence between LTE-U and Wi-Fi systems on unlicensed spectrum is realized.

7.2 Future Research

In this section, we discuss three future research directions to our most interest.

7.2.1 Cross-Layer CC Selection and (De)activation for CA-Based LTE-A systems

According to Section 2.2, the dynamic CC management should consist of two levels. The first level is the semi-dynamic CC selection/configuration, which determines the optimal subset of CCs for a user to transmit on when the connection is established or reconfigured. It usually takes place several minutes or even an hour apart. The other level is the dynamic CC (de)activation function, which decides whether a UE should sleep on a particular assigned CC in each TTI. It takes place every 1ms and can significantly reduce the power consumption for the user terminal. In most existing literature, the second level is usually not considered. Therefore, cross-layer strategies must be designed to make full use of the two-time-scale dynamic CC management to optimize the target performance while saving as much power as possible.

In such strategies, the CC (de)activation is carried out through MAC signalings after the layer-3 CC configuration process and only the configured CCs can be (de)activated. In each TTI, the optimal CC subset that should be deactivated is determined for each user. For the best-effort users, the objective is to maximize the average user throughput; for the users with minimal throughput requirements, the objective is to maximize the total/average power saving while satisfying the QoS constraints. To determine the optimal subsets, possible thresholds can be derived. The users which exceed higher threshold can be assigned with more CCs. The considered decision factors should involve the instantaneous cell load conditions, the channel statistics of the wireless fadings, the service outrage probabilities, etc.

7.2.2 Interference Management in "Green" HetNets with CA

Although dense deployment of HetNets has been recognized as a desirable and feasible solution for increasing the network capacity and QoS provisioning, a large number of SBSs account for high capital expenditure (CAPEX) and operational expenditure (OPEX). In ad-

dition, a high volume of energy is required to power multiple SBSs in the cellular network, which also results in more greenhouse gases and carbon footprints, besides the increased energy costs. Therefore, one of the most critical issues in the next generation 5G wireless network is to design and develop new green technologies to reduce the energy consumption and costs of HetNets.

To this end, we intend to investigate the RRM problems in the context of green HetNets with CA, where the MBSs are still powered by electricity grid while the small cells rely fully or partially on the renewable energy sources such as solar and wind. But unlike traditional energy supply, green energy sources are by nature sustainable in the long term yet are variable and unstable in the short term, as the energy harvesting process is highly dependent on the charging environment, i.e., the weather, the geo-location, and the time. Thus, the primary goal of RRM in green HetNets transits from optimizing the network performance to optimizing the network sustainability. Analytical models need to be developed to characterize the charging dynamics of the SBSs and resource allocation strategies need to be designed based on the charging statistics.

7.2.3 Traffic Balancing between Licensed Spectrum to Unlicensed spectrum

The LTE-U traffic balancing is subject to the unlicensed regulations thus presenting unique features. Traffic balancing in LTE-U context should take into consideration the user activities from other independent unlicensed systems in order to protect their performance. Due to LBT features, one LTE-U small cell may not be able to occupy the unlicensed spectrum for a certain period even if it is needed by LTE-U users. Hence, the LTE-U performance in unlicensed bands are time-varying and heavily dependent on other systems channel access activities. Consequently, a dilemma arises: on one hand, LTE-U small cell tends to assign more users to the unlicensed spectrum to reduce interference to the macrocell users; on the other hand, user performance on the unlicensed spectrum varies a lot, thus making it hard to

provision QoS guarantee. Therefore, tradeoffs need to be made therein to provide LTE users with optimized traffic dispatch over different bands.

Chapter 8

Related Publications

8.1 Books and Book Chapters

1. M. Wang, **R. Zhang**, and X. Shen, *Mobile Electric Vehicles: Online Charging and Discharging*, SpringerBriefs in Electrical and Computer Engineering, Springer Verlag, 2015. (ISBN-10: 3319251287, ISBN-13: 978-3319251288)

2. **R. Zhang**, M. Wang, L. X. Cai, Y. Cheng, X. Shen, and L. Xie, "Resource Management in Sustainable Green HetNets with Renewable Energy Sources," 5G Mobile Communications, Springer.

8.2 Journal Papers

1. **R. Zhang**, M. Wang, X. Shen, and L. Xie, "Probabilistic Analysis on QoS Provisioning for Internet of Things in LTE-A Heterogeneous Networks with Partial Spectrum Usage," *IEEE Internet of Things Journal*, Oct. 2015.

2. **R. Zhang**, M. Wang, L. X. Cai, Z. Zheng, X. Shen, and L. Xie, "LTE-Unlicensed: The Future of Spectrum Aggregation for Cellular Networks," *IEEE Wireless Communications Magazine*, vol. 22, no. 3, pp. 150-159, 2015.

3. **R. Zhang**, Z. Zheng, M. Wang, X. Shen, and L. Xie, "Equivalent Capacity in Carrier Aggregation-Based LTE-A Systems: A Probabilistic Analysis," *IEEE Trans. on Wireless Communications*, vol. 13, no. 11, pp. 6444-6460, Nov. 2014.

4. M. Wang, Q. Shen, **R. Zhang**, H. Liang, and X. Shen, "Vehicle-Density based Adaptive MAC for High Throughput in Drive-thru Networks," *IEEE Internet of Things Journal*, vol 1, no. 6, pp. 533-543, Dec. 2014.

5. M. Wang, H. Shan, T. H. Luan, N. Lu, **R. Zhang**, X. Shen, and F. Bai, "Asymptotic Throughput Capacity Analysis of VANETs Exploiting Mobility Diversity," *IEEE Trans. on Vehicular Technology*, 2015.

6. A. T. Gamage, H. Liang, **R. Zhang**, and X. Shen, "Device-to-Device Communication Underlaying Converged Heterogeneous Networks," *IEEE Wireless Communications Magazine*, vol. 21, no. 6, pp. 98-107, Dec. 2014.

7. Z. Zheng, X. Zhang, L. X. Cai, **R. Zhang**, and X. Shen, "Sustainable Communication and Networking in Two-tier Green Cellular Networks," *IEEE Wireless Communications*, vol. 21, no. 4, pp. 47-53, Aug. 2014.

8. M. Wang, H. Shan, R. Lu, **R. Zhang**, X. Shen, and F. Bai, "Real-Time Path Planning Based on Hybrid-VANET-Enhanced Transportation System," *IEEE Trans. on Vehicular Technology*, vol. 64, no. 5, pp. 1664-1678, 2015.

9. M. Wang, H. Liang, **R. Zhang**, R. Deng, and X. Shen, "Mobility-Aware Coordinated Charging for Electric Vehicles in VANET-Enhanced Smart Grid," *IEEE J. Selected Areas of Communications*, vol. 32, no. 7, pp. 1344-1360, July 2014.

10. Z. Zheng, L. X. Cai, **R. Zhang**, and X. Shen, "RNP-SA: Joint Relay Placement and Sub-carrier Allocation in Wireless Communication Networks with Sustainable Energy," *IEEE Trans. on Wireless Communications*, vol. 11, no. 10, pp. 3818-3828, 2012.

8.3 Conference Papers

1. **R. Zhang**, M. Wang, L. X. Cai, X. Shen, L. Xie, and Y. Cheng, "Modeling and analysis of MAC protocol for LTE-U co-existing with Wi-Fi," *IEEE Globecom'15*, 2015.

2. **R. Zhang**, M. Wang, Z. Zheng, X. Shen, and L. Xie, "Stochastic Geometric Performance Analysis for Carrier Aggregation in LTE-A Systems," *IEEE ICC'14*, 2014.

3. **R. Zhang**, Z. Zheng, M. Wang, X. Shen, and L. Xie, "Equivalent Capacity Analysis of LTE-Advanced Systems with Carrier Aggregation", *IEEE ICC'13*, 2013.

4. **R. Zhang**, M. Wang, Z. Zheng, X. Shen, and L. Xie, "Cross-Layer Carrier Selection and Power Control for LTE-A Uplink with Carrier Aggregation", *IEEE Globecom'13*, 2013.

5. Y. Chen, Y. Wang, **R. Zhang**, and X. Shen, "Optimization on Power Splitting Ratio Design for K-tier HCNs with Opportunistic Energy Harvesting," *IEEE ICC'15*, 2015.

6. Z. Zheng, L. X. Cai, N. Zhang, **R. Zhang**, and X. Shen, "A Game Theoretical Approach for Energy Trading in Wireless Networks Powered by Green Energy," *IEEE Globecom'14*, 2014.

7. M. Wang, M. Ismail, **R. Zhang**, X. Shen, E. Serpedin, and K. A. Qaraqe, "A Semi-distributed V2V Fast Charging Strategy based on Price Control," *IEEE Globecom'14*, 2014. (**Best Paper Award**)

8. M. Wang, H. Liang, R. Deng, **R. Zhang** and X. Shen, "VANET Based Online Charging Strategy for Electric Vehicles," *Proc. IEEE Globecom'13*, 2013.

Bibliography

[1] ITU-R M.2133, *Requirements, evaluation criteria and submission templates for the Development of IMT-Advanced*, 2008.

[2] 3GPP RP-080137, *Proposed SID on LTE-advanced*, NTT DoCoMo, etc., 3GPP TSG RAN meeting No. 39.

[3] D. Astély, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "LTE: the evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, no 4, pp. 44-51, Apr. 2009.

[4] G. Yuan et al., "Carrier aggregation for LTE-advanced mobile communication systems, *IEEE Communication Magazine*, vol. 48, no. 2, pp. 88-93, Feb. 2010.

[5] M. Iwamura et al., "Carrier aggregation framework in 3GPP LTE-advanced, *IEEE Communication Magazine*, vol. 48, no. 8, pp. 60-67, Aug. 2010.

[6] K. I. Pedersen et al., "Carrier aggregation for LTE-advanced: Functionality and performance aspects, *IEEE Communication Magaine*, vol. 49, no. 6, pp. 89-95, Jun. 2011.

[7] H. Kuwano and M. Kojima, *Mobile communication system, radio network controller, radio terminal, data delivering method, and program for the method*, US Patent App. 10/747,958.

[8] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: Deployment

scenarios and operational challenges," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 148-155, Feb. 2012.

[9] http://www.visionmobile.com/blog/2007/12/do-we-really-need-femto-cells/.

[10] P. Bhat, S. Nagata, L. Campoy, I. Berberana, T. Derham, G. Liu, X. Shen, P. Zong, and J. Yang, "LTE-advanced: An operator perspective," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 104-114, 2012.

[11] 3GPP TR 36.808 v10.0.0, *Evolved universal terrestrial radio access (E-UTRA); carrier aggregation; base station (BS) radio transmission and reception (Release 10).* Tech. Spec. Group Radio Access Network, Jun. 2012.

[12] 3GPP, Tech. Spec. 36.304, v10.3.0, *Evolved universal terrestrial radio access (E-UTRA); User equipment (UE) procedures in idle mode*, 2011.

[13] C. Ciochina and H. Sari, "A review of OFDMA and single-carrier FDMA", *Wireless Conference (EW), 2010 European*, pp. 706-710, 2010.

[14] Z. Shen, A. Papasakellariou, J. Montojo, D. Gerstenberger, and F. Xu, "Overview of 3GPP LTE-advanced carrier aggregation for 4G wireless communications," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 122-130, Feb. 2012.

[15] S. Yang and Y. Lin, "Modeling UMTS discontinuous reception mechanism," *IEEE Trans. on Wireless Communications*, vol. 4, no. 1, pp. 312-319, 2005.

[16] 3GPP TS 36.321, v10.1.0, *Evolved universal terrestrial radio access (E-UTRA); Medium access control (MAC) protocol specification (Release 10)*, TSG RAN, 2011.

[17] 3GPP TS 36.104 v9.1.0, *Base station (BS) radio transmission and reception (release 9).* Tech. Spec. Group Radio Access Network, Sept. 2009.

[18] G. Yuan, X. Zhang, W. Wang, and Y. Yang, "Carrier aggregation for LTE-advanced mobile communication systems," *IEEE Communications Magazine*, vol. 48, no. 2, pp. 88-93, Feb. 2010.

[19] K. I. Pedersen, F. Frederiksen, C. Rosa, H. Nguyen, L. G. U. Garcia, and Y. Wang, "Carrier aggregation for LTE-advanced: functionality and performance aspects," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 89-95, Jun. 2011.

[20] Y. Wang, K. I. Pedersen, T. B. Sørensen, and P. E. Mogensen, "Carrier load balancing and packet scheduling for multi-carrier systems," *IEEE Trans. on Wireless Communications*, vol. 9, no. 5, pp. 1780-1789, May 2010.

[21] M. Awad, V. Mahinthan, M. Mehrjoo, X. Shen, and J. W. Mark, "A dual-decomposition-based resource allocation for OFDMA networks with imperfect CSI," *IEEE Trans. on Vehicular Technology*, vol. 59, no. 5, pp. 2394-2403, Jun. 2010.

[22] J. Mao, G. Xie, J. Gao, and Y. Liu, "Energy efficiency optimization optimization for OFDM-based cognitive radio systems: A water-filling factor aided search method," *IEEE Trans. on Wireless Communications*, vol. 12, no. 5, pp. 2366-2375, May 2013.

[23] M. S. Alam, J. W. Mark, and X. Shen, "Relay selection and resource allocation for multi-user cooperative OFDMA networks," *IEEE Trans. on Wireless Communications*, vol. 12, no. 5, pp. 2193-2205, May 2013.

[24] M. Bohge, J. Gross, A. Wolisz, and M. Meyer, "Dynamic resource allocation in OFDM systems: An overview of cross-layer optimization principles and techniques," *Network, IEEE*, vol. 21, no. 1, pp 53-59, Feb. 2007.

[25] S. Chieochan and E. Hossain, "Adaptive radio resource allocation in OFDMA systems: A survey of the state-of-the-art approaches," *Wireless Communications and Mobile Computing*, vol. 9, no. 4, pp 513-527, Apr. 2009.

[26] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y. Kim, E. Kim, and Y. Cheong, "An overview of radio resource management in relay-enhanced OFDMA-based networks," *Communications Surveys and Tutorials, IEEE*, vol. 12, no. 3, pp 422-438, 2010.

[27] D. López-Pérez, X. Chu, A. V. Vasilakos, and H. Claussen, "On distributed and coordinated resource allocation for interference mitigation in self-organizing LTE networks," *IEEE Trans. on Networking*, vol. 21, no. 4, pp. 1145-1158, Aug. 2013.

[28] J. Brown and J. Y. Khan, "Key performance aspects of an LTE FDD based smart grid communications network," *Computer Communications*, vol. 36, no. 5, pp. 551-561, Mar. 2013.

[29] K. I. Pedersen, T. E. Kolding, F. Frederiksen, I. Z. Kovács, D. Laselva, and P. E. Mogensen, "An overview of downlink radio resource management for UTRAN long-term evolution," *IEEE Communications Magzine*, vol. 47, no. 7, pp. 86-93, Jul. 2009.

[30] M. Kalil, A. Shami, A. Al-Dweik and S. Muhaidat, "Low-complexity power-efficient schedulers for LTE uplink with delay-sensitive traffic," *IEEE Trans. on Vehicular Technology*, vol. 64, no. 10, pp. 4551-4564, Oct. 2015.

[31] M. Kalil, A. Shami, and A. Al-Dweik, "QoS-aware power-efficient scheduler for LTE uplink," *IEEE Trans. on Mobile Computing*, vol. 14, no. 8, pp. 1672-1685, Aug. 2015.

[32] R. Zhang, M. Wang, Z. Zheng, X. Shen, and L. Xie, "Stochastic geometric performance analysis for carrier aggregation in LTE-A systems," *Proc. IEEE ICC'14*, Sydney, Australia, June 10-14, 2014.

[33] R. Zhang, Z. Zheng, M. Wang, X. Shen, and L. Xie, "Equivalent capacity analysis of LTE-advanced systems with carrier aggregation," *Proc. IEEE ICC'13*, Budapest, Hungary, Jun. 2013.

[34] R. Zhang, Z. Zheng, M. Wang, X. Shen, and L. Xie, "Equivalent capacity in carrier aggregation-based LTE-A systems: A probabilistic analysis," *IEEE Trans. on Wireless Communications*, vol. 13, no. 11, pp. 6444-6460, Nov. 2014.

[35] R. Zhang, M. Wang, Z. Zheng, X. Shen, and L. Xie, "Cross-layer carrier selection and power control for LTE-A uplink with carrier aggregation", *Proc. IEEE Globecom'13*, Atlanta, GA, USA, Dec. 9-13, 2013.

[36] F. Wu, Y. Mao, X. Huang, and S. Leng, "A joint resource allocation scheme for OFDMA-based wireless networks with carrier aggregation," *Proc. IEEE WCNC'12*, Shanghai, China, Apr. 2012.

[37] F. Liu, W. Xiang, Y. Zhang, K. Zheng, and H. Zhao, "A novel QoE-based carrier scheduling scheme in LTE-Advanced networks with multi-service," *Proc. IEEE VTC Fall'12*, Quebec City, Canada, Sept. 2012.

[38] C. Li, B. Wang, W. Wang, Y. Zhang, and X. Chang, "Component carrier selection for LTE-A systems in diverse coverage carrier aggregation scenario," *Proc. IEEE PIMRC'12*, Sydney, Austrilia, Sept. 2012.

[39] A. Abrardo, M. Belleschi, P. Detti, and M. Moretti, "Message passing resource allocation for the uplink of multi-carrier multi-format systems," *IEEE Trans. on Wireless Communications*, vol. 11, no. 1, pp. 130-141, 2012.

[40] F. Liu, K. Zheng, W. Xiang, and H. Zhao, "Design and performance analysis of an energy-efficient uplink carrier aggregation scheme," *IEEE Journal on Selected Areas in Communications*, Vol. 32, No. 12, pp. 1-11, May 2013.

[41] H. Wang, H. Nguyen, C. Rosa, and K. Pedersen, "Uplink multi-cluster scheduling with MU-MIMO for LTE-Advanced with carrier aggregation," *Proc. IEEE WCNC'12*, Shanghai, China, Apr. 2012.

[42] G. Yuan, X. Zhang, W. Wang, and Y. Yang, "Carrier aggregation for LTE-advanced mobile communication systems," *IEEE Communications Magazine*, vol. 48, No. 2, pp. 88-93, 2010.

[43] Y. Rui, P. Cheng, M. Li, Q. T. Zhang, and M. Guizani, "Carrier aggregation for LTE-Advanced: Uplink multiple access and transmission enhancement features," *IEEE Wireless Communications*, Vol. 20, No. 4, pp. 101-108, Aug. 2013.

[44] C. Wang and T. Lin, "Method of handling component carrier activation and deactivation and communication device thereof," *US Patent App. 13/071,519*, Mar. 25, 2011.

[45] C. Wu and B. P. Sebire, "Discontinuous reception for carrier aggregation," *US Patent App. 14/115,884*, Aug. 12, 2011.

[46] R. Prakash, J. M. Damnjanovic, P. Gaal, P. A. Agashe, V. A. Gheorghiu, M. Kitazoe, and R. Palanki, "Multiple carrier activation/deactivation in wireless communications," *US Patent App. 12/953,356*, Nov. 23, 2010.

[47] F. Yin, "An application aware discontinuous reception mechanism in LTE-advanced with carrier aggregation consideration," *Annals of telecommunications-annales des télécommunications*, vo. 67, no. 3-4, pp. 147-159, 2012.

[48] C. Zhong, T. Yang, L. Zhang, and J. Wang, "A new discontinuous reception (DRX) scheme for LTE-advanced carrier aggregation systems with multiple services," *Proc. IEEE VTC Fall'2011*, 2011.

[49] 3GPP R4-091910, *LTE-A MC RF requirements for contiguous carriers*, May 2009.

[50] H. Wang, C. Rosa, and K. I. Pedersen, "Uplink component carrier selection for LTE-advanced systems with carrier aggregation," *Proc. IEEE ICC'11*, Kyoto, Japan, Jun. 2011.

[51] Y. Tseng and C. Huang, "Analysis of femto base station network deployment," *IEEE Trans. on Vehicular Technology*, vol. 61, No. 2, pp. 748-757, 2012.

[52] J. Yun and K. G. Shin, "Adaptive interference management of OFDMA femtocells for co-Channel deployment," *IEEE Journal on Selected Areas in Communications*, vol. 29, No. 6, pp. 1225-1241, 2011.

[53] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, No. 9, pp. 1479-1489, 2010.

[54] L. G. U. Garcia, K. I. Pedersen, and P. E. Mogensen, "Autonomous component carrier selection: Interference management in local area environments for LTE-advanced", *IEEE Communications Magazine*, vol. 47, No. 9, pp. 110-116, Sept. 2009.

[55] B. Wang, Y. Zhang, W. Wang, M. Lei, and L. Jiang, "A cooperative downlink power setting scheme for CA-based femtocells," *Proc. IEEE VTC Spring'12*, Yokohama, Japan, May 2012.

[56] H. Wang, C. Rosa, and K. I. Pedersen, "Dedicated carrier deployment in heterogeneous networks with inter-site carrier aggregation," *Proc. IEEE WCNC'13*, Shanghai, China, Apr. 2013.

[57] L. Duan, J. Huang, and B. Shou, "Economics of femtocell service provision," *IEEE Trans. on Mobile Computing,* vol. 12, no. 11, pp. 2261-2273, Nov. 2013.

[58] S. Bu, F. R. Yu, and Y. Qian, "Energy-efficient cognitive heterogeneous networks powered by the smart grid," *Proc. IEEE INFOCOM'13*, Turin, Italy, Apr. 2013.

[59] L. G. U. Garcia, I. Z. Kovcs, K. I. Pedersen, G. W. O. Costa, and P. E. Mogensen, "Autonomous component carrier selection for 4G femtocells - a fresh look at an old problem," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 525-537, 2012.

[60] L. G. U. Garcia, F. Sanchez-Moya, J. Villalba-Espinosa, K. I. Pedersen, and P. E. Mogensen, "Enhanced uplink carrier aggregation for LTE-Advanced femtocells," *Proc. IEEE VTC Fall'11*, San Francisco, USA, Sept. 2011.

[61] F. Sanchez-Moya, J. Villalba-Espinosa, L. G. U. Garcia, K. I. Pedersen, and P. E. Mogensen, "On the impact of explicit uplink information on autonomous component carrier selection for LTE-A femtocells," *Proc. IEEE VTC Spring'11*, Yokohama, Japan, May 2011.

[62] F. Hu, K. Zheng, L. Lei, and W. Wang, "A distributed inter-cell interference coordination scheme between femtocells in LTE-Advanced networks," *Proc. IEEE VTC Spring'11*, Yokohama, Japan, May 2011.

[63] I. Akyildiz, W. Y. Lee, M. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magzine*, vol. 46, No. 4, pp. 40-48, 2008.

[64] A. Galindo-Serrano, L. Guipponi, and M. Dohler, "Cognition and docition in OFDMA-based femtocell networks," *IEEE Proc. GLOBECOM10*, Barcelona, Spain, Dec. 2010.

[65] R. Zhang, M. Wang, X. Shen, L. Xie, "Probabilistic analysis on QoS provisioning for internet of things in LTE-A heterogeneous networks with partial spectrum usage," *IEEE Internet of Things Journal*, Oct. 2015.

[66] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1029-1046, 2009.

[67] Y. Zhong and W. Zhang, "Multi-channel hybrid access femtocells: A stochastic geometric analysis," *IEEE Trans. on Communications*, vol. 61, no. 7, pp. 3016-3026, 2013.

[68] A. Bleicher, "A surge in small cell sites," *IEEE Spectrum*, vol. 50, no. 1, Jan. 2013, pp. 38-39.

[69] 3GPP RP-140808, "Review of regulatory requirements for unlicensed spectrum," Alcatel-Lucent, Alcatel-Lucent Shanghai Bell, Ericsson, Huawei, HiSilicon, IAESI, LG, Nokia, NSN, Qualcomm, NTT Docomo, June 2014.

[70] Nokia, "Nokia LTE for unlicensed spectrum," June 2014.

[71] T. Nihtilä, V. Tykhomyrov, O. Alanen, M. A. Uusitalo, A. Sorri, M. Moisio, S. Iraji, R. Ratasuk, and N. Mangalvedhe, "System performance of LTE and IEEE 802.11 co-existing on a shared frequency band," *Proc. IEEE WCNC'13*, 2013.

[72] Qualcomm, "Qualcomm research LTE in unlicensed spectrum: Harmonious coexistence with Wi-Fi," 2014.

[73] 3GPP RP-140808, "Review of regulatory requirements for unlicensed spectrum," 2014.

[74] F. Liu, E. Bala, E. Erkip, M. C. Beluri, and R. Yang, "Small cell traffic balancing over licensed and unlicensed bands," *IEEE Trans. on Vehicular Tech.*, 2015.

[75] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen, and L. Xie, "LTE-Unlicensed: The future of spectrum aggregation for cellular networks," *IEEE Wireless Communications Magazine*, vol. 22, no. 3, pp. 150-159, 2015.

[76] R. Zhang, M. Wang, L. X. Cai, X. Shen, L. Xie, and Y. Cheng, "Modeling and analysis of MAC protocol for LTE-U co-existing with Wi-Fi," *IEEE Globecom 2015*.

[77] R. Guérin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 7, Sept. 1991.

[78] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. on Wireless Communicaitons*, vol. 2, no. 4, pp. 630-643, Jul. 2003.

[79] R. E. Miles, "On the homogeneous planar Poisson point process," *Mathematical Biosciences*, vol. 6, pp. 85-127, 1970.

[80] J.-S. Ferenc and Z. Néda, "On the size distribution of Poisson Voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518-526, 2007.

[81] P. T. Brady, "A model for generating on-off speech patterns in two-way conversations," *Bell System Tech. J.*, 48(Sept. 1969): 2445-2472.

[82] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Trans. on Communications*, vol. 36, no. 7, pp. 834-844, July 1988.

[83] B. Błaszczyszyn, M. K. Karray, and H.-P. Keeler, "Using poisson processes to model lattice cellular networks," *Proc. IEEE INFOCOM'13*, Turin, Italy, Apr. 2013.

[84] A. Goldsmith, *Wireless communications.* Cambridge university press, pp. 24-84, 2005.

[85] C. C. M. Grinstead and J. L. Snell, *Introduction to probability.* American Mathematical Soc., pp. 291-305, 1997.

[86] M. Schwartz, *Broadband integrated networks.* Prentice Hall PTR, vol. 19, pp. 26-29 and 126-140, 1996.

[87] G. E. P. Box, J. S. Hunter, and W. G. Hunter, *Statistics for experimenters: Design, innovation, and discovery, 2nd edition.* Wiley Online Library, pp. 130, May 2005.

[88] Y. Nesterov, A. Nemirovskii, and Y. Ye, *Interior-point polynomial algorithms in convex programming*, SIAM, vol. 13, 1994.

[89] O. Järv, R. Ahas, E. Saluveer, B. Derudder, and F. Witlox, "Mobile phones in a traffic flow: A gegographical perspective to evening rush hour traffic analysis using call detail records," *PLoS ONE*, vol. 7, no. 11, Nov. 2012.

[90] S. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, "Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink," *Proc. IEEE INFOCOM'09*, Rio de Janeiro, Brazil, Apr. 2009.

[91] 3GPP TS 36.213 v8.1.0, *Evolved universal terrestrial radio access (E-UTRA). Physical layer procedures (Release 8)*, Tech. Spec. Group Radio Access Network, Nov. 2007.

[92] D. J. Dechene and A. Shami, "Energy-aware resource allocation strategies for LTE uplink with synchronous HARQ constraints," *IEEE Trans. on Mobile Computing*, vol. 13, no. 2, pp. 422-433, Feb. 2014.

[93] D. J. Dechene and A. Shami, "Energy efficient resource allocation in SC-FDMA uplink with synchronous HARQ constraints," *Proc. IEEE ICC'11*, Kyoto, Japan, 2011.

[94] H. Wang, C. Rosa, and K. I. Pedersen, "Performance of uplink carrier aggregation in LTE-advanced systems," *Proc. IEEE VTC Fall'10*, Ottawa, Canada, Sept. 2010.

[95] 3G TR 25.943 v0.1.0, *Deployment aspects*, Tech. Spec. Group Radio Access Network, Feb. 2000.

[96] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *INFORMS Operations Research*, vol. 14, no. 4, pp. 699-719, 1966.

[97] K. R. Malekshan, W. Zhuang, and Y. Lostanlen, "An energy efficient MAC protocol for fully connected wireless ad hoc networks," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 10, pp. 5729-5740, 2014.

[98] R. Murias, M. Beluri, and M. Rudolph, "LTE-U co-existence mechanisms," *IEEE 802.19-14/0035r2*, 2014.

[99] L. X. Cai, X. Shen, J. W. Mark, L. Cai, Y. Xiao, "Voice capacity analysis of WLAN with unbalanced traffic," *IEEE Trans. on Vehicular Tech.*, vol. 55, no. 3, pp. 752 - 761, 2006.

[100] M. F. Usman, A. Hussain, and F. Nadeem, “Saturation throughput analysis of IEEE 802.11e EDCA through analytical model,” *International Journal of Wireless Information Networks,* vol. 21, no. 2, pp. 101-113, 2014.

Vita

Ran Zhang received his B.E. degree on Electronic Information Science and Technology from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2010. He is now working towards his Ph.D. degree with Broadband Communication Research (BBCR) Group from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, under the supervision of Professor Xuemin (Sherman) Shen and Professor Liang-liang Xie.

His current research interests include radio resource management in 4G/5G mobile communication networks (with emphasis on heterogeneous networks (HetNets), carrier aggregation (CA) in LTE-A standard, and LTE-Unlicensed technology), wireless green networks, and electrical vehicle charging control in smart grids.