Coexistence of Wireless Networks for Shared Spectrum Access

Bo Gao

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Engineering

Jung-Min Park, Co-Chair
Yaling Yang, Co-Chair
Y. Thomas Hou
Sandeep K. Shukla
Danfeng Yao

September 11, 2014
Blacksburg, Virginia

Keywords: Opportunistic Spectrum Access, Cognitive Radio Networks, White Space Networks, Shared Spectrum Access, Spectrum Sharing, Network Coexistence

Coexistence of Wireless Networks for Shared Spectrum Access

Bo Gao

(ABSTRACT)

The radio frequency spectrum is not being efficiently utilized partly due to the current policy of allocating the frequency bands to specific services and users. In opportunistic spectrum access (OSA), the "white spaces" that are not occupied by primary users (a.k.a. incumbent users) can be opportunistically utilized by secondary users. To achieve this, we need to solve two problems: (i) primary-secondary incumbent protection, i.e., prevention of harmful interference from secondary users to primary users; (ii) secondary-secondary network coexistence, i.e., mitigation of mutual interference among secondary users. The first problem has been addressed by spectrum sensing techniques in cognitive radio (CR) networks and geolocation database services in database-driven spectrum sharing. The second problem is the main focus of this dissertation. To obtain a clear picture of coexistence issues, we propose a taxonomy of heterogeneous coexistence mechanisms for shared spectrum access. Based on the taxonomy, we choose to focus on four typical coexistence scenarios in this dissertation.

Firstly, we study sensing-based OSA, when secondary users are capable of employing the channel aggregation technique. However, channel aggregation is not always beneficial due to dynamic spectrum availability and limited radio capability. We propose a channel usage model to analyze the impact of both primary and secondary user behaviors on the efficiency of channel aggregation. Our simulation results show that user demands in both the frequency and time domains should be carefully chosen to minimize expected cumulative delay.

Secondly, we study the coexistence of homogeneous CR networks, termed as self-coexistence, when co-channel networks do not rely on inter-network coordination. We propose an uplink soft frequency reuse technique to enable globally power-efficient and locally fair spectrum sharing. We frame the self-coexistence problem as a non-cooperative game, and design a local heuristic algorithm that achieves the Nash equilibrium in a distributed manner. Our

simulation results show that the proposed technique is mostly near-optimal and improves self-coexistence in spectrum utilization, power consumption, and intra-cell fairness.

Thirdly, we study the coexistence of heterogeneous CR networks, when co-channel networks use different air interface standards. We propose a credit-token-based spectrum etiquette framework that enables spectrum sharing via inter-network coordination. Specifically, we propose a game-auction coexistence framework, and prove that the framework is stable. Our simulation results show that the proposed framework always converges to a near-optimal distributed solution and improves coexistence fairness and spectrum utilization.

Fourthly, we study database-driven OSA, when secondary users are mobile. The use of geo-location databases is inadequate in supporting location-aided spectrum sharing if the users are mobile. We propose a probabilistic coexistence framework that supports mobile users by locally adapting their location uncertainty levels in order to find an appropriate trade-off between interference mitigation effectiveness and location update cost. Our simulation results show that the proposed framework can determine and adapt the database query intervals of mobile users to achieve near-optimal interference mitigation with minimal location updates.

This work was supported by National Science Foundation (NSF) under grants CNS-0746925, CNS-0831865, CNS-0910531, CNS-1054697, CNS-1228903, CNS-1314598, and IIP-1265886, and Institute for Critical Technology and Applied Science (ICTAS) at Virginia Tech.

To my beloved family

— my daughter Cheryl T. Gao, my wife Junfei Tang,

and my parents Futing Gao, Zhihui Yang

Acknowledgments

Over the past five years, I have received so much support, generosity, and friendship from so many people. I would first like to thank my advisors Dr. Yaling Yang and Dr. Jung-Min Park. I could not have made it without their guidance and expertise. I thought I was not smart enough, but they gave me the confidence to pursue an academic career. I would also like to thank my advisory committee members Dr. Thomas Hou, Dr. Sandeep Shukla, and Dr. Danfeng Yao for their efforts and contributions to this work. I would give special thanks to Dr. Thomas Hou for his courses that helped me better understand my research area.

I would like to thank my colleagues Kaigui Bian, Zhenhua Feng, Chuan Han, Ting Wang, Jingyao Zhang, Yi Tang, Behnam Bahrak, Zhenhe Pan, Chang Liu, Vireshwar Kumar, Seungmo Kim, Sudeep Bhattarai, Xiangwei Zheng, Kexiong Zeng, Yanzi Dou, He Li, and Jinshan Liu from the SHINE and ARIAS labs. Many ideas were inspired by the discussions with them. Special thanks go to Kaigui Bian for his experience in achieving academic success.

I would like to thank all my friends, who made my life in Blacksburg colorful, pleasant, and memorable. Being far away from home is not easy, but they made me feel at home.

I would give my heartfelt thanks to my wife Junfei Tang for her love, understanding, care, and trust. Our daughter Cheryl is the best gift I have ever received. They made me realize how beautiful my life could be. They gave me the courage to move on. I would also thank my parents Futing Gao and Zhihui Yang for everything they did for me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the motivation, challenges, and contributions of this dissertation.

1.1 Motivation

In accordance with the current fixed spectrum allocation policy, radio frequency bands have been allocated to diverse specific services and users by spectrum regulators, such as the Federal Communications Commission (FCC) in the United States and the Office of Communications (Ofcom) in the United Kingdom. In general, wireless spectrum is divided into two classes: licensed spectrum and unlicensed spectrum. On the one hand, unlicensed spectrum bands such as the industrial, scientific, and medical (ISM) bands (i.e., 902-928 MHz, 2.400-2.500 GHz, and 5.725-5.875 GHz) have become congested due to the popularity of personal wireless devices. On the other hand, considerable licensed spectrum bands such as the broadcast digital TV bands (i.e., 54-698 MHz) have not been fully utilized [1, 2, 3]. To redress the imbalance between congested unlicensed spectrum and underutilized licensed spectrum, the concept of opportunistic spectrum access (OSA) has been proposed. In OSA, primary users (a.k.a. incumbent users) and secondary users are usually co-located and share

the same swaths of spectrum. The fallow spectrum that is not occupied by primary users is called "white spaces", and can be opportunistically utilized by secondary users. To achieve this, we need to solve two spectrum sharing problems: (i) *primary-secondary incumbent protection*, i.e., prevention of harmful interference from secondary users to primary users; (ii) *secondary-secondary network coexistence*, i.e., mitigation of mutual interference among secondary users. The first problem has been addressed by spectrum sensing techniques in cognitive radio (CR) networks and geolocation database services in database-driven spectrum sharing [4, 5, 6]. A reconfigurable CR can dynamically identify white spaces through various incumbent detection techniques and adjust its operating parameters following the changes in spectrum availability in order to ensure incumbent protection. However, only relying on CRs for sensing-based OSA is not sufficient to achieve reliable protection of incumbents and efficient utilization of white spaces, since spectrum sensing techniques are either ineffective or expensive for the discovery of low-power or passive primary users. For this reason, the use of incumbent geolocation databases for database-driven OSA has found favor with the spectrum regulators and the wireless industry. A geolocation database has access to the detailed operating information of primary users, such as service types, channel reservations, and protection requirements. Based on the up-to-date incumbent characteristics, the database is able to make use of radio wave propagation models to compute location-specific spectrum availability. A secondary user can directly query the database and retrieve the information regarding available white spaces at a given location. In both the sensing-based and database-driven OSA, the second problem for shared spectrum access has received little attention in the literature, and thus is the main focus of this dissertation.

The problems of secondary-secondary coexistence in OSA mainly deal with the mitigation of mutual interference among co-located secondary networks. Interference mitigation mechanisms are usually in the physical (PHY) or media access control (MAC) layers of a wireless network system. In the PHY layer, a variety of techniques can be used to suppress harmful interference, such as interference cancellation, smart antennas, and transmit power control. However, these PHY-layer techniques are not sufficient to effectively maximize spectrum

utilization. In the MAC layer, coexisting networks can operate in separate channels (in the frequency domain) or time slots (in the time domain), or at different locations (in the space domain). In this dissertation, resource allocation problems in the MAC layer are addressed, which directly determine whether white spaces can be efficiently utilized.

1.2 Challenges and Contributions

As the technology for shared spectrum access evolves, an increasing number of coexistence mechanisms are proposed, such as the case in TV white spaces (TVWSs) [7]. Several emerging standards have prescribed wireless access technologies operating in the TVWSs, and more standards are expected to follow suit. The coexistence of heterogeneous CR networks in TVWSs has become a challenging problem that we have to face. To offer a clear picture of coexistence issues and related technical challenges, and moreover, shed light on the possible solution space, we propose a taxonomy of heterogeneous coexistence mechanisms [8]. According to the taxonomy, this dissertation covers all four types of coexistence architectures. First, spectrum access can be guided by probabilistic modeling of user behaviors and prediction of white space opportunities. A channel usage model is developed to tell how much white spaces, free of primary users and coexisting secondary users, are available on average under certain user arrival and departure processes. Second, shared spectrum access can be achieved by an autonomous mechanism. Neither coexistence infrastructures nor coordination channels are required. Coexisting networks rely on local observations to make spectrum access decisions. Third, shared spectrum access can be achieved by a coordinated mechanism. Coexisting networks coordinate spectrum access over coordination channels, but each network makes decisions for itself without the need for extra coexistence infrastructures. Fourth, shared spectrum access can be achieved by a database-driven centralized mechanism. Coexisting networks are registered with a geolocation database, and a central spectrum manager optimizes spectrum sharing among the registered networks.

1.2.1 Sensing-Based OSA

Firstly, in the early stage of CR-related research, it is necessary to find out the impact of both primary and secondary user behaviors on the availability of white space opportunities, especially when secondary users are capable of employing the channel aggregation technique. A CR is able to identify white space channels through spectrum sensing and aggregate multiple channels at a time through channel aggregation. In regular wireless networks, channel aggregation can achieve capacity gain according to the Shannon's theorem. In CR networks, however, it may not always be beneficial due to dynamic spectrum availability and limited radio capability. The availability of white spaces is affected not only by primary users but also by co-channel secondary users when coexistence issue is taken into account. However, the existing channel usage models neglect the impact of secondary user activity that is needed for coexistence analysis. Moreover, the common assumption in these models that channel idle/busy duration is exponentially distributed is not true according to recent measurement study [9]. Hence, we aim to propose a channel usage model that is general to capture both primary and secondary user behaviors. Using this model to probabilistically characterize white space availability, we want to study the strategy for performing channel aggregation in the sense of average performance.

In the first scenario, we study the efficiency of channel aggregation in sensing-based OSA in consideration of dynamic spectrum availability and limited radio capability [10]. In order to support channel aggregation with low chance of causing co-channel interference, we propose a channel usage model, which is used to analyze the impact of both primary and secondary user behaviors on white space availability. This model is general to capture a wide range of user behaviors. We model the delay costs for performing channel aggregation. Then, a channel aggregation strategy is derived to minimize expected cumulative delay. Our numerical and simulation results based on real data of primary user activity show that user demands in both the frequency and time domains should be carefully chosen in reality.

1.2.2 Coexistence of Homogeneous CR Networks

Secondly, as license-exempt CR networks become possible, the coexistence of homogeneous CR networks is a common issue, especially when the coexisting networks belong to different network operators, i.e., inter-network coordination is not available. Multiple co-located networks can share the same swaths of spectrum through spatial spectrum reuse. However, even though we consider the coexistence of homogeneous CR networks, termed as self-coexistence, it is still challenging to achieve efficient spectrum reuse in CR networks, especially in the uplink. Unlike most licensed networks such as cellular networks that are deployed with careful central frequency planning, there is typically no a central entity that can coordinate spectrum sharing among unlicensed networks. It is very likely that co-located, co-channel CR networks are managed by different network operators, such as the cases of IEEE 802.16h [11] and IEEE 802.22 [12]. In most situations, reliable and real-time inter-network coordination cannot be available. Furthermore, inter-network interference in the uplink caused by user-generated signals is much more dynamic and unpredictable than that in the downlink caused by relatively static broadcast signals from base stations. Hence, we aim to propose a resource allocation mechanism for co-channel self-coexistence of CR networks in the uplink, when the coexisting networks do not rely on inter-network coordination.

In the second scenario, we propose an uplink soft frequency reuse technique to enable globally power-efficient and locally fair spectrum sharing among homogeneous CR networks [13, 14]. We frame the self-coexistence problem as a non-cooperative game. In each network cell, uplink resource allocation (URA) problem is decoupled into two subproblems: subchannel allocation (SCA) and transmit power control (TPC). We provide a unique optimal solution for TPC, while present a low-complexity heuristic for SCA. We study multi-cell URA as a game, and design a local heuristic algorithm that achieves the Nash equilibrium in a distributed manner. Our simulation results in both multi-operator and single-operator coexistence scenarios show that the proposed technique is mostly near-optimal and improves self-coexistence in spectrum utilization, power consumption, and intra-cell fairness.

1.2.3 Coexistence of Heterogeneous CR Networks

Thirdly, as more and more wireless standards take advantage of CR technology to offer new services, the coexistence of heterogeneous CR networks becomes a major issue, especially when a central coexistence infrastructure is not available. The coexisting networks can be heterogeneous, i.e., use different air interface standards, such as the case in TVWSs [7]. A number of emerging standards have prescribed wireless access technologies operating in the TVWSs, and more standards are expected to follow suit. When we consider the coexistence of heterogeneous CR networks, inter-network coordination is needed for certain coexistence etiquette or a coexistence infrastructure is needed for centralized resource allocation. Recently, IEEE 802.19.1 [15] and a COGEU project [16] are being developed to provide general infrastructure-based solutions to the coexistence of 802 or non-802 networks. However, we are more interested in the case without a central coexistence infrastructure. Hence, we aim to propose a spectrum sharing mechanism for co-channel coexistence of heterogeneous CR networks, which rely on inter-network coordination to share white spaces.

In the third scenario, we propose a credit-token-based spectrum etiquette framework to enable spectrum sharing among heterogeneous CR networks with equal priority [17]. Specifically, we propose a game-auction coexistence framework. Each network acts as either an offerer or a requester, and coexists with other networks via a non-cooperative game and a truthful multi-winner auction. The framework addresses the trade-offs among social welfare and offerer's revenue in the auction and requester's utility in the game. We prove that the framework guarantees system stability. Our simulation results show that the proposed coexistence framework always converges to a near-optimal distributed solution and improves coexistence fairness and spectrum utilization.

1.2.4 Database-Driven OSA

Fourthly, as the geolocation database technology evolves, database-driven OSA begins to take the place of sensing-based OSA as the major spectrum sharing architecture, and location-aided spectrum sharing becomes a new challenge, especially when secondary users are mobile. Lately, the FCC has issued a NPRM for enabling small cell use in the 3.5 GHz band [18, 19]. In the proposed framework, spectrum access system (SAS) serves as a central spectrum manager, which incorporates a geolocation database and various interference mitigation techniques. A secondary user updates the database with its location information, so that the updated knowledgebase can facilitate location-aided spectrum sharing. However, secondary user mobility results in location uncertainty at SAS-like database server, which may harm white space availability and spectrum reuse efficiency. When supporting mobile users, each user can query and update the database at a certain interval [19]. On the one hand, during these intervals, the database may have to rely on past location information to ensure incumbent protection and support network coexistence. This leads to considerable loss of white space opportunities and increased likelihood of inaccurate resource allocations. On the other hand, it can be unacceptable to blindly reduce the intervals because of limited database processing capacity, significant communication overhead rise, and needless network resource waste. This causes the database to become a bottleneck in processing spectrum sharing information. Hence, we aim to propose a practical strategy for adapting the database query intervals to strike an appropriate trade-off between the effectiveness of interference mitigation and the cost of location updates.

In the fourth scenario, we propose a probabilistic coexistence framework to support mobile users in database-driven OSA [20]. The framework incorporates the solutions to two core problems: white space allocation (WSA) at the database and location update control (LUC) at the users. For WSA, we derive a centralized real-time solution to mitigate mutual interference among secondary users and protect primary users against harmful interference. For LUC, we design a local two-level strategy to enable both movement-driven and interference-

driven control of location uncertainty. This strategy makes an appropriate trade-off between interference mitigation effectiveness and location update cost. Our simulation results in both model-driven and trace-driven simulation experiments show that the proposed coexistence framework can determine and adapt the database query intervals of mobile users to achieve near-optimal interference mitigation with minimal location updates.

1.3 Organization

The remainder of this dissertation is organized as follows.

- Chapter 2 introduces the preliminary knowledge of the coexistence of wireless networks — a taxonomy of heterogeneous coexistence mechanisms.
- Chapter 3 studies sensing-based OSA, when secondary users are capable of employing channel aggregation technique.
- Chapter 4 studies the coexistence of homogeneous CR networks, when co-channel networks do not rely on inter-network coordination.
- Chapter 5 studies the coexistence of heterogeneous CR networks, when co-channel networks use different air interface standards.
- Chapter 6 studies database-driven OSA, when secondary users are mobile.
- Chapter 7 draws conclusions.
- Chapter 8 discusses future work, which addresses the integration of spectrum sensing into database-driven spectrum sharing.

Chapter 2

Preliminaries

This chapter introduces the preliminary knowledge of the coexistence of wireless networks. We propose a taxonomy of heterogeneous coexistence mechanisms for shared spectrum access in white space spectrum [8]. This taxonomy aims to offer a clear picture of coexistence issues and related technical challenges, and moreover, shed light on the possible solution space. Heterogeneous coexistence mechanisms are also applied to self-coexistence scenarios.

2.1 Overview of Coexistence Mechanisms

In 2008, the FCC authorized license-exempt operation of CR devices in digital TV bands (i.e., VHF/UHF 54-698 MHz) to further utilize the fallow spectrum in these bands, which is commonly termed as TVWSs. As the most representative white space spectrum, TVWSs have received great attention from the wireless community due to the desirable signal propagation characteristics in the TV bands and the overcrowding in the ISM bands. A number of emerging standards have prescribed wireless access technologies operating in the TVWSs, and more standards are expected to follow suit. These heterogeneous secondary networks have equal priority, and are expected to coexist in the same TV bands. Recently, the FCC

is opening up 3.5 GHz band for new OSA opportunities. The proposed three-tier spectrum sharing framework in the 3.5 GHz band requires heterogeneous database-driven networks that may not have equal priority to coexist safely. In these types of white space spectrum, the coexistence of dissimilar wireless networks necessitates the need for heterogeneous coexistence (HC) mechanisms [7, 15] to mitigate harmful mutual interference among the coexisting networks. Different from HC issues in the ISM bands, it is more challenging to support HC in OSA due to the signal propagation characteristics in licensed bands, the spectrum agility of systems operating in white space spectrum, and the disparity of PHY/MAC strategies of the coexisting systems. The excellent RF penetration depth of signals in TV bands can cause high level of inter-network interference, while the federal incumbents in 3.5 GHz band are very sensitive to harmful interference from lower-tier users. Furthermore, the heterogeneity of coexisting networks in terms of network architecture, radio access technologies, and required quality of service (QoS) poses challenging problems in the design of HC mechanisms.

At this point, we briefly review the recently published or emerging wireless standards that prescribe license-exempt operation in white space spectrum. In TVWSs, traditional wireless networks ranging from short-range personal or local area networks to long-range metropolitan area networks have their counterparts that support OSA. In December 2009, ECMA-392 [21] was finalized as the first standard for personal/portable CR devices operating in TVWSs. It specifies MAC and PHY operations and defines several self-coexistence mechanisms for inter-network coordination and interference mitigation. In 2008, Google and Microsoft proposed the idea of WiFi-like operation in TVWSs, called WiFi 2.0 or WhiteFi [22]. A new standard based on this idea was formalized as IEEE 802.11af [23], which targets higher rate and wider coverage than the current WiFi services by using CR-enabled access points (APs) and user terminals. Besides incumbent protection, IEEE 802.11af also needs to address the coexistence of co-located APs, even though the coexistence mechanisms are yet to be finalized. License-exempt operation of existing licensed networks, such as LTE and IEEE 802.16, further creates new challenges. At present, license-exempt LTE [24] is still in its infancy, but IEEE 802.16h [11] was published as a standard amendment for license-exempt WiMAX in July

2010. The cognitive WiMAX proposes various uncoordinated and coordinated coexistence mechanisms, which can be suitable for the coexistence of metropolitan area networks with heterogeneous others in TVWSs. In July 2011, IEEE 802.22 [12] was released as a new standard for regional area networks located in rural areas using TVWSs. Like ECMA-392, several self-coexistence mechanisms are defined in it to mitigate mutual interference among co-located networks belonging to different operators. Besides specific network systems, IEEE 802.19.1 [15] and COGEU project [16] are being developed to provide general solutions to the coexistence of 802 or non-802 networks in various CR-enabled use cases—e.g., campus, apartment complex, and home. A typical IEEE 802.19.1 system consists of a coexistence manager, which acts as a centralized resource allocator, and a coexistence enabler, which acts as a coexistence information collector maintaining interfaces between the coexistence manager and coexisting CR networks. As the FCC issued a NPRM for enabling OSA in 3.5 GHz band [18, 19] lately, a total of 150 MHz spectrum that is being used by some critical federal and non-federal incumbents, such as radar and satellite services, has become a new focus of spectrum regulators and researchers. In the proposed framework, spectrum sharing can be achieved among three tiers of users, including tier-1 incumbent access users, tier-2 priority access users, and tier-3 general authorized access users. The services of tier-1 and tier-2 users should be protected from harmful interference in certain exclusion zones, while tier-3 users have to opportunistically access the band. All the users need to register in the SAS, which incorporates a geolocation database and various interference mitigation techniques. The SAS serving as a central spectrum manager is able to collect the location information of registered secondary users and utilize the information to achieve database-driven protection of primary users and coexistence of secondary users.

The HC issues are also common in ISM bands. A widely studied coexistence scenario is the co-channel coexistence of low-power IEEE 802.15 networks and high-power IEEE 802.11 networks [25, 26, 27]. This problem is also addressed by IEEE 802.15.2. Another popular coexistence scenario is the co-channel coexistence of contention-based IEEE 802.11/802.15 networks and reservation-based IEEE 802.16 networks [28, 29].

Table 2.1: List of coexistence mechanisms.

Coexistence Mechanism	Sources
Adaptive modulation and coding	802.11, LTE, 802.16, 802.22
Centralized coexistence framework	[15, 16, 30]
Coexistence beacon signaling	ECMA-392, 802.22
Coexistence control channel	[29], 802.16h
Coexistence frame scheduling	802.16h
Coexistence information database	[7, 15, 18, 19], 802.16h
Cognitive pilot channel	[7]
Collocated coexistence messaging	[31], 802.16
Cooperative busy tone signaling	[27]
Coordinated contention-based protocol	802.16h
Credit-token-based coexistence protocol	802.16h
Dynamic frequency/channel selection	[15, 24], ECMA-392, 802.16h, 802.22
Interference cancellation and suppression	[28], 802.15.2
Internet-server-facilitated messaging	[15]
Listen before talk	[15], ECMA-392, 802.11, 802.16h
Opportunistic channel access	[25, 26, 28]
Smart antenna	[7, 15], ECMA-392, 802.11, LTE, 802.16, 802.22
Time/frequency-division multiple access	[7, 15], 802.15.2
Transmit power control	[7, 15, 32], ECMA-392, 802.11, LTE, 802.16, 802.22

The coexistence mechanisms defined in the literature are summarized in Table 2.1, and will be introduced as examples in the next section in detail.

2.2 Taxonomy of Coexistence Mechanisms

The proposed taxonomy classifies coexistence mechanisms, especially HC mechanisms in TVWSs, using a diverse set of criteria as shown in Figure 2.1. For each category of the taxonomy, we provide examples and discuss the pros and cons of related mechanisms. Note that some of the mechanisms discussed in the following discussions are not applicable to HC exclusively—they are also applicable to self-coexistence scenarios. In Table 2.2, we map the coexistence mechanisms summarized in Table 2.1 to our taxonomy.

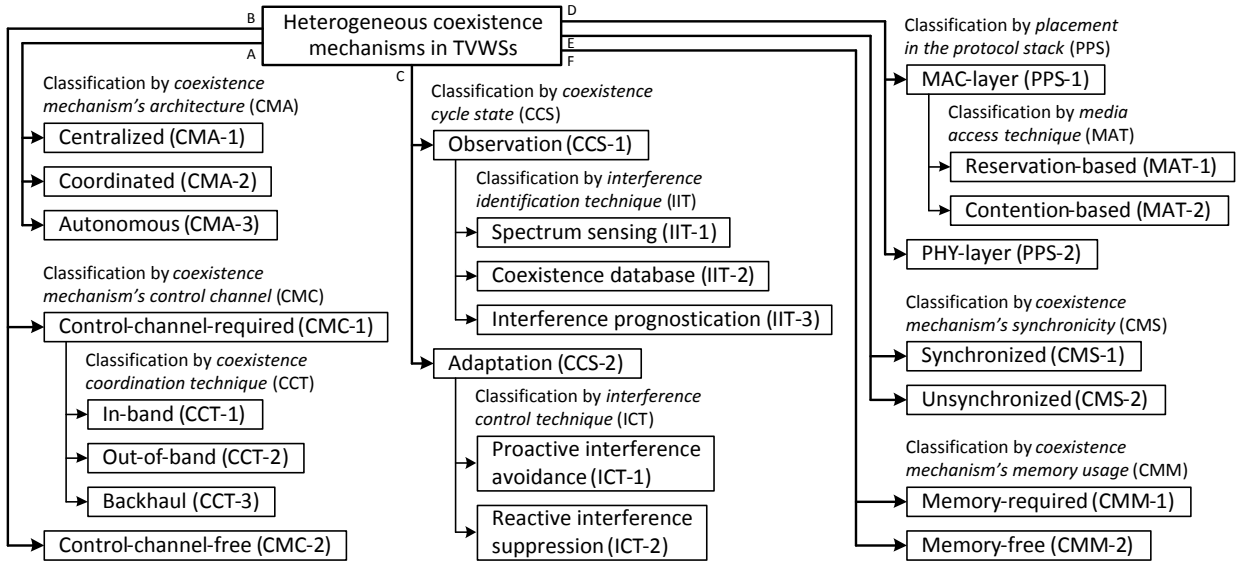

Figure 2.1: A taxonomy of coexistence mechanisms.

2.2.1 Coexistence Mechanism's Architecture

Spectrum sharing among coexisting networks can be achieved in several ways depending on whether or not decision-making coexistence infrastructures and inter-network coordination channels are available. Based on the coexistence mechanism's architecture (CMA), HC mechanisms are classified into *centralized*, *coordinated*, and *autonomous* categories.

CMA-1: Centralized Mechanisms

The operation of centralized mechanisms has to rely on both coexistence infrastructures and coordination channels. A coexistence scenario for centralized mechanisms is illustrated in Figure 2.2.a. Each operator can deploy a single or multiple infrastructures to carry out centralized spectrum sharing via a star-topology or a cluster-based architecture. In each cluster, a coexistence infrastructure minimizes mutual interference among registered coexisting networks based on centrally collected coexistence information. Furthermore, multiple coexistence infrastructures can coordinate with each other for further collaboration. For example, the standard-independent centralized coexistence framework in IEEE 802.19.1 defines a coexistence manager for central management of all associated networks. In addition, the

collaboration of multiple 802.19.1 coexistence managers or cluster-head equipments [30] can be supported through certain inter-cluster coordination interfaces. The SAS in 3.5 GHz band is another example. The centralized mechanisms can operate independently of existing wireless standards, and thus, registered networks can be heterogeneous and no major standard modifications are required. They can effectively minimize inter-network interference by utilizing centrally collected global coexistence information. However, centralized solutions incur costly new infrastructures and subsequent operational costs. The effectiveness of centralized mechanisms is diminished when a significant fraction of the coexisting devices/networks are not registered and are not under central control.

CMA-2: Coordinated Mechanisms

The operation of coordinated mechanisms does not require coexistence infrastructures but does need coordination channels. A coexistence scenario for coordinated mechanisms is illustrated in Figure 2.2.b. The coexisting networks coordinate with each other through various information signaling and retrieving techniques. Based on the collected coexistence information, each network can make decisions to mitigate inter-network interference. For example, the cooperative busy tone signaling technique [27] helps high-power IEEE 802.11 networks detect signaling messages from co-channel, low-power IEEE 802.15.4 networks to avoid 802.11 devices' domination over 802.15.4 devices in the channel contention process. The coordinated mechanisms achieve effective inter-network interference mitigation. However, they are often limited to certain coexistence scenarios. More details about coordination channels will be discussed in the next subsection.

CMA-3: Autonomous Mechanisms

Each coexisting network has to make use of autonomous mechanisms to achieve best-effort interference mitigation when neither coexistence infrastructures nor coordination channels are available. A coexistence scenario for autonomous mechanisms is illustrated in Figure 2.2.c. Each network performs resource allocation and manages inter-network interference only based on local observations. For example, the dynamic frequency/channel selection

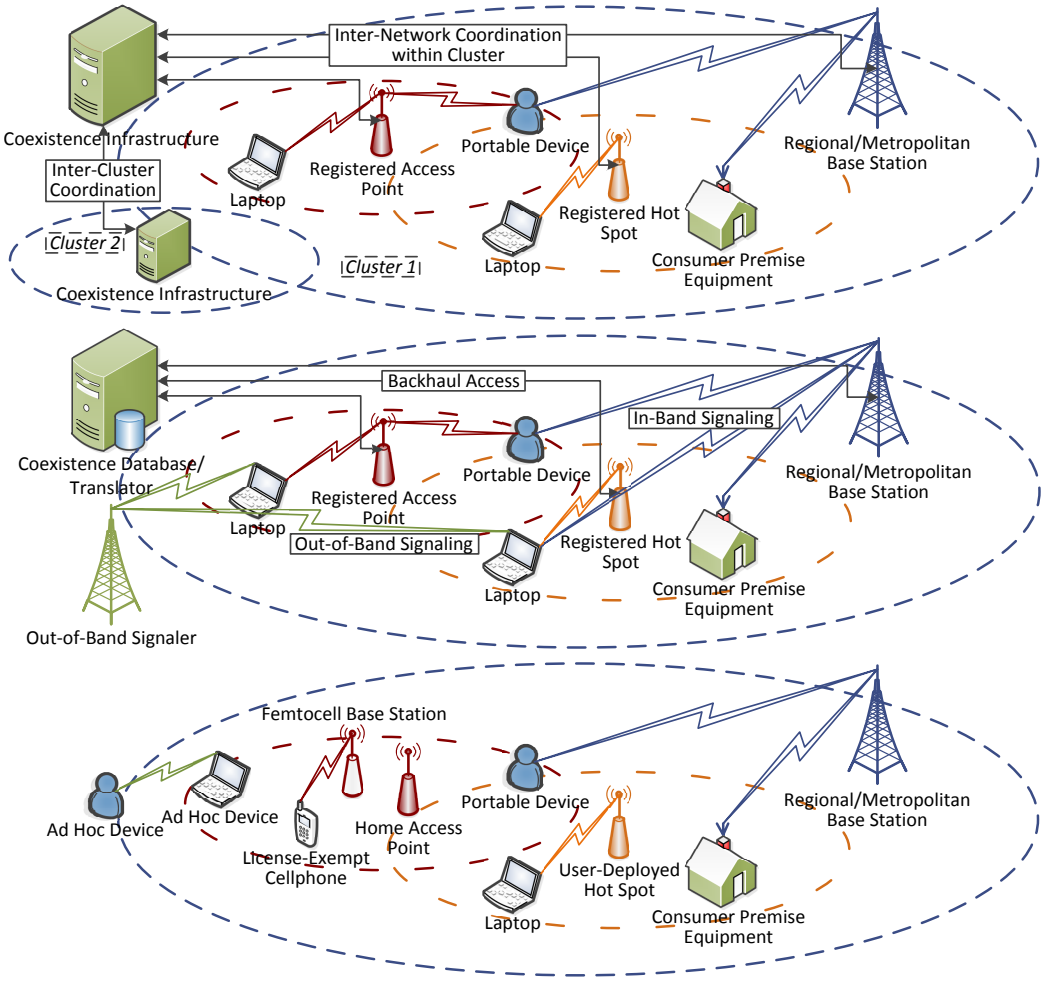

Figure 2.2: Examples of heterogeneous coexistence scenarios: a) centralized mechanisms (top); b) coordinated mechanisms (middle); c) autonomous mechanisms (bottom).

technique allows each network to select or switch to the channel with the least amount of interference according to the local evaluation of channel quality. The listen before talk policy prescribes a device to access spectrum based on the outcome of local spectrum sensing. The autonomous mechanisms are low in complexity and are easy to adapt to dynamic environments. They can be integrated with centralized and coordinated mechanisms to build hybrid mechanisms. However, autonomous mechanisms by themselves may not sufficiently mitigate inter-network interference due to their best-effort nature.

Table 2.2: Mapping of coexistence mechanisms to classification methods.

	CMA-1	CMA-2	CMA-3	CMC-1: CCT-1	CMC-1: CCT-2	CMC-1: CCT-3	CMC-2	CCS-1: IIT-1	CCS-1: IIT-2	CCS-1: IIT-3	CCS-2: ICT-1	CCS-2: ICT-2	PPS-1: MAT-1	PPS-1: MAT-2	PPS-2	CMS-1	CMS-2	CMM-1	CMM-2
Adaptive modulation and coding			✓				✓					✓			✓		✓		✓
Centralized coexistence framework	✓					✓					✓		✓			✓		✓	
Coexistence beacon signaling		✓		✓				✓					✓			✓			✓
Coexistence control channel		✓			✓			✓					✓			✓			✓
Coexistence frame scheduling		✓			✓			✓			✓		✓			✓			✓
Coexistence information database		✓				✓			✓				✓			✓		✓	
Cognitive pilot channel		✓			✓			✓					✓			✓		✓	
Collocated coexistence messaging		✓		✓				✓					✓			✓			✓
Cooperative busy tone signaling		✓		✓				✓						✓			✓		✓
Coordinated contention-based protocol		✓			✓			✓			✓			✓		✓			✓
Credit-token-based coexistence protocol		✓			✓			✓			✓		✓			✓			✓
Dynamic frequency/channel selection			✓				✓	✓			✓		✓				✓	✓	
Interference cancellation and suppression			✓				✓					✓			✓	✓		✓	
Internet-server-facilitated messaging		✓				✓		✓					✓			✓		✓	
Listen before talk			✓				✓	✓			✓			✓			✓		✓
Opportunistic channel access			✓				✓			✓				✓			✓	✓	
Smart antenna			✓				✓					✓			✓		✓		✓
Time/frequency-division multiple access	✓					✓					✓		✓			✓		✓	
Transmit power control			✓				✓					✓			✓		✓		✓

2.2.2 Coexistence Mechanism's Control Channel

The availability of control channels for inter-network coordination determines the design of HC mechanisms. Based on the coexistence mechanism's control channel (CMC), HC mechanisms are classified into *control-channel-required* and *control-channel-free* categories.

CMC-1: Control-Channel-Required Mechanisms

Both the centralized mechanisms (CMA-1) and coordinated mechanisms (CMA-2), whose operations necessarily require inter-network coordination, fall into this classification. Based on the coexistence coordination technique (CCT), control-channel-required mechanisms are further classified into *in-band*, *out-of-band*, and *backhaul* categories.

CMC-1: CCT-1: In-Band Mechanisms

To deliver coexistence information to coexisting neighbors, each network can periodically broadcast coexistence signaling messages on its data channels, such as in-band signaling illustrated in Figure 2.2.b. For example, the coexistence beacon signaling in ECMA-392 and IEEE 802.22 defines specific time slots in each network's regular superframes for periodic broadcast of coexistence beacons. Although such a technique is used for self-coexistence purpose, it can be a HC mechanism when coexisting networks can decode signaling messages between them. The collocated coexistence messaging technique [31] lets each multi-radio user terminal forward in-band signaling messages among its multiple home networks. The in-band mechanisms do not require extra infrastructures or control channels for inter-network coordination. However, they only work when coexisting networks use the same radio access technology or when heterogeneous devices can decode each others' signaling messages.

CMC-1: CCT-2: Out-of-Band Mechanisms

Instead of relying on data channels, coexisting networks can broadcast coexistence signaling messages on a dedicated or dynamically established common control channel, such as out-of-band signaling illustrated in Figure 2.2.b. For example, the coexistence control channel in IEEE 802.16h supports secondary synchronization, user detection, interference evaluation,

and inter-system communication. The cognitive pilot channel [7] always carries up-to-date coexistence information, broadcasted by operators or third-party entities, which can be retrieved by each network on demand. The out-of-band mechanisms can be used in direct inter-network negotiations if a standardized signaling message format is adopted. However, they fully rely on the existence and reliability of a common control channel.

CMC-1: CCT-3: Backhaul Mechanisms

If available, wired backhaul links can be used to coordinate coexisting networks. As shown in Figure 2.2.b, each network can either access a central coexistence database or coordinate with neighbors using the method of message translation. For example, the coexistence information database [18, 19], which is implemented on servers connected to the Internet, provides on-demand lookup functionality. The Internet-server-facilitated messaging technique [15] enables inter-network coordination over the Internet by providing message translation and forwarding services. The backhaul mechanisms can provide relatively complete knowledge of the spectrum environment in the vicinity of each network. They enable active reporting and retrieving of coexistence information instead of passive network detection, so that the complexity of each device/network is reduced. However, backhaul links may not cover all the coexisting networks, and thus the coexistence information can be incomplete.

CMC-2: Control-Channel-Free Mechanisms

In the context of our taxonomy, control-channel-free mechanisms are classified under a category equivalent to the one that the autonomous mechanisms (CMA-3) are classified under.

2.2.3 Coexistence Cycle State

The simplified cognition cycle of a CR consists of three states: observation, decision, and adaptation [33]. Based on this cycle, we propose a classification of HC mechanisms based on the coexistence cycle state (CCS), which includes: *observation* and *adaptation* categories.

CCS-1: Observation Mechanisms

The purpose of the observation state is to identify the presence or even the source of inter-network interference. Based on the interference identification technique (IIT), observation mechanisms are further classified into *spectrum sensing*, *coexistence database*, and *interference prognostication* categories.

CCS-1: IIT-1: Spectrum Sensing Mechanisms

For the detection of coexisting networks, each network can either locally scan white space spectrum via spectrum sensing or cooperatively exchange coexistence information with neighbors through inter-network coordination channels. Note that the purpose of spectrum sensing here is to detect coexisting secondary networks instead of primary incumbents. For example, the scanning-based opportunistic channel access technique [25] enables each IEEE 802.15.4 network to scan multiple channels potentially interfered by IEEE 802.11 networks, and keep searching for a better channel selection by using simulated annealing optimization method. The metric of channel quality is computed in terms of the energy of detected 802.11 interference and the number of heard 802.15.4 beacons. The coexistence beacon signaling and coexistence control channel techniques facilitate cooperative spectrum sensing and coexistence information exchange by providing means for in-band or out-of-band inter-network coordination. The spectrum sensing mechanisms are relatively easy to implement. However, they are unreliable in real-world coexistence scenarios, and need to be augmented with other methods for reliable detection of coexisting networks.

CCS-1: IIT-2: Coexistence Database Mechanisms

A coexistence database storing geolocation and operation information of secondary networks can be utilized by each coexisting network to help identify potential sources of mutual interference. For example, the coexistence information database in IEEE 802.16h stores the shared information regarding the actual and intended usage of spectrum resource for certain local regions. Besides base stations, user terminals also contribute to complete the database by providing interference information pertinent to themselves. The coexistence database

mechanisms provide a more practical and effective means of detecting coexisting networks. However, they cannot detect the presence of unregistered devices/networks at the database authority, and the coexistence information may not be updated in real time.

CCS-1: IIT-3: Interference Prognostication Mechanisms

The past spectrum sensing results can help each network make short-term predictions on spectrum availability and potential interference. Interference prognostication techniques include modeling and machine learning. These techniques can be used to predict the behaviors of potential interferers by leveraging their past spectrum access behaviors. For example, the learning-based opportunistic channel access techniques [25, 26] enable each IEEE 802.15.4 network to learn the statistical regularity of IEEE 802.11 operation and predict the "white spaces" free of 802.11 traffic. The ideas are similar to the modeling of primary user activity in order to empirically prevent secondary users from using the same spectrum being occupied by primary users. The interference prognostication mechanisms offer low-power networks more opportunities of spectrum access by circumventing high-power networks. However, their effectiveness highly relies on the accuracy of predictions, which is very challenging to guarantee for the operation in white space spectrum.

CCS-2: Adaptation Mechanisms

Once inter-network interference has been detected, a network can adapt to its interference environment by taking various measures to change its transmission characteristics. Based on the interference control technique (ICT), adaptation mechanisms are further classified into *proactive interference avoidance* and *reactive interference suppression* categories.

CCS-2: ICT-1: Proactive Interference Avoidance Mechanisms

Upon detecting harmful interference from neighbors, each network can choose to directly switch to a better-quality channel, if such a channel is available, or wait until the channel becomes vacant. If coordination channels are available, inter-network interference can even be avoided in advance. For example, the dynamic frequency/channel selection technique

provides each network with a list of candidate channels that can be used for channel switching whenever needed. The time/frequency-division multiple access is a common interference avoidance technique that centrally enables coexisting networks to operate in separate time slots or channels. The proactive interference avoidance mechanisms are suitable for the coexistence of CR networks, since CR devices are capable of dynamic spectrum access. However, they necessarily require the support of observation mechanisms to identify the candidate channels that are free of interference.

CCS-2: ICT-2: Reactive Interference Suppression Mechanisms

Instead of trying to avoid interference from neighbors, interference suppression techniques are used to directly alleviate or suppress the experienced interference. For example, the interference cancellation and suppression technique [28] allows any network that is being interfered with to utilize adaptive filters or prior knowledge of interferers to estimate and cancel the interference step by step. The transmit power control policy can be used to mitigate co-channel or adjacent-channel interference when the coexisting networks are managed by the same operator. The reactive interference suppression mechanisms help to support non-exclusive co-channel spectrum sharing as long as inter-network interference is tolerable. However, their effectiveness is usually limited, especially when the interference comes from high-power interferers that belong to different operators.

2.2.4 Placement in Protocol Stack

In general cases, inter-network interference is mitigated in the MAC or PHY layer of the protocol stack. Based on the placement in the protocol stack (PPS), HC mechanisms are classified into *MAC-layer* and *PHY-layer* categories.

PPS-1: MAC-Layer Mechanisms

Most HC mechanisms are placed in the MAC layer. Each network can choose to either reserve or contend for spectrum access. Based on the media access technique (MAT), MAC-layer mechanisms are further classified into *reservation-based* and *contention-based* categories.

PPS-1: MAT-1: Reservation-Based Mechanisms

The reservation of spectrum resources can be made in the time, frequency, or space domains. Note that space-division mechanisms are included in the PPS-2 category. In the time domain, multiple co-channel networks can take turns to access the shared channel in separate time frames or slots. For example, the coexistence frame scheduling in IEEE 802.16h divides time frames into Master, Slave, and Shared subframes, which can be scheduled by each base station for uplink and downlink in a flexible mode. The operation of a Master system in its Master subframes should be protected from harmful interference caused by concurrent Slave systems, and coexisting networks equally share the role of Master system on a rotating basis. In Shared subframes, all the coexisting networks may operate in parallel under the limits on transmit power levels. In the frequency domain, multiple coexisting networks can simultaneously access the same spectrum but use separate channels or subchannels by direct or orthogonal spectrum splitting. Furthermore, time-frequency resource blocks may be conceived for greater flexibility and granularity in spectrum sharing. For example, the credit-token-based coexistence protocol in IEEE 802.16h permits auction-based spectrum leasing among coexisting networks for channel reservation in the subsequent time frames. Each network can be either an offerer or a requester to transfer the ownership of spectrum dynamically. The reservation-based mechanisms can guarantee coexistence fairness and reliable throughput of coexisting networks regardless of the discrepancy in their channel definitions, signal characteristics, or transmit power levels. However, they need to be supported by coordination channels or even extra coexistence infrastructures.

PPS-1: MAT-2: Contention-Based Mechanisms

The contention of spectrum resources is another way of managing media access. For example, the coordinated contention-based protocol in IEEE 802.16h prescribes the networks with reservation-based MAC to periodically halt transmissions so that the resulting idle time frames can be utilized by the networks with contention-based MAC. The contention-based mechanisms are easy to implement, and do not require strict inter-network synchronization. However, they do not always guarantee coexistence fairness and constantly reliable spectrum access due to the randomness of contention results.

PPS-2: PHY-Layer Mechanisms

In the PHY layer, a variety of techniques can be used to mitigate inter-network interference. For example, the smart antennas can be used to reduce the coexistence-related interference by minimizing sidelobe radiation. The spatial reuse of shared spectrum can be improved by directional interference patterns. The adaptive modulation and coding technique enhances coexistence via dynamically adaptable PHY parameters according to the varying radio environments, such as path loss and interference. Most PHY-layer mechanisms also belong to the autonomous mechanisms (CMA-3), so further discussion is neglected.

2.2.5 Coexistence Mechanism's Synchronicity

The mitigation of mutual interference can be facilitated by the mechanisms synchronized across coexisting networks. Based on the coexistence mechanism's synchronicity (CMS), HC mechanisms are classified into *synchronized* and *unsynchronized* categories.

CMS-1: Synchronized Mechanisms

Many coexistence techniques rely on accurate inter-network synchronization, in either MAC or PHY layer. For example, the coordinated contention-based protocol requires the coexisting networks with reservation-based MAC to synchronize their quiet periods so that the carrier sense of the networks with contention-based MAC can discover such opportunities.

The coexistence beacon signaling and coexistence control channel techniques enable faster network detection if the networks synchronize with each other, due to the periodicity of in-band or out-of-band signaling. The synchronized mechanisms address coexistence issues through precise separation of coexisting networks in the time domain. However, accurate synchronization is hard to implement without extra infrastructures to support.

CMS-2: Unsynchronized Mechanisms

Inter-network synchronization is not necessary for a number of coexistence techniques. For example, the listen before talk policy enables coexisting networks to contend for spectrum access in an asynchronous manner. The cooperative busy tone signaling technique serves as an enhancement of carrier sense multiple access (CSMA) protocol for IEEE 802.15.4 networks, and does not need synchronization with co-channel IEEE 802.11 networks that use CSMA as well. The unsynchronized mechanisms can be readily implemented. However, in most cases, they only alleviate inter-network interference but cannot avoid it.

2.2.6 Coexistence Mechanism's Memory Usage

The repository of coexistence information can be helpful sometimes to mitigate inter-network interference. Based on the coexistence mechanism's memory usage (CMM), HC mechanisms are classified into *memory-required* and *memory-free* categories.

CMM-1: Memory-Required Mechanisms

In some coexistence techniques, a memory is needed to store coexistence information that is useful for spectrum access. For example, the coexistence information database maintains up-to-date geolocation and operation information of coexisting networks. The machine-learning-based opportunistic channel access needs to record recent history of spectrum sensing results and maintain a knowledgebase to make reasonable interference predictions. The memory-required mecha- nisms can help make faster and more thoughtful decisions. However, the costs for memory consumption can be high, especially in large-scale complex coexistence scenarios, e.g., an apartment building in a dense urban area.

CMM-2: Memory-Free Mechanisms

There are still a lot of coexistence techniques that do not require memory usage or only require a negligible size of memory. For example, the collocated coexistence messaging technique for each multi-radio user terminal directly forwards signaling messages from one of its home networks to another one without the need for recording the forwarded messages. The cooperative busy tone signaling technique only requires IEEE 802.15.4 signalers to emit busy tones to IEEE 802.11 receivers for spectrum reservation. The listen before talk policy is another typical example. The memory-free mechanisms are suitable for autonomous networks that do not have much system resources. However, their achieved performance in terms of coexistence fairness, spectrum utilization, or throughput may not be as good as that by memory-required mechanisms due to limited coexistence knowledge.

Chapter 3

Channel Aggregation in Cognitive Radio Networks

In this chapter, we focus on the first scenario to study sensing-based OSA, when secondary users are capable of employing the channel aggregation technique [10].

3.1 Challenges and Contributions

In CR networks, incumbent protection is the major concern. When secondary users (SUs) opportunistically access the identified white spaces in licensed spectrum, existing primary users (PUs) should always be protected against harmful interference. Desiring more transmission capacity, a SU equipped with CR is able to transmit over wider bandwidth, since capacity is proportional to bandwidth according to the Shannon's theorem. However, it may not be true in OSA. For incumbent protection, wider bandwidth for a SU leads to higher risk of interfering with PUs or being interrupted by returning PUs. Furthermore, if network coexistence is considered, wider bandwidth for a SU also results in higher chance of causing

co-channel interference with other coexisting SUs. Hence, increasing bandwidth may not always be beneficial for capacity expansion in OSA.

Usually, licensed spectrum is divided into a number of discrete channels. Hence, efficient spectrum utilization can be achieved by properly enabling each SU to access multiple white space channels at a time [34]. We borrow the definition of channel aggregation (CA) from the draft version of IEEE 802.22 [35] to study the feature of assembling noncontiguous channels for communication. Technically, CA can be implemented based on orthogonal frequency division multiplexing (OFDM) [2, 36, 37] or multiple radios [38]. However, because CR technology itself is not mature yet, the complexity of implementing CA is still very high.

In this chapter, we are interested in studying the efficiency of CA in consideration of dynamic spectrum availability and limited radio capability. We make the following contributions.

- We propose a channel usage model to investigate the impact of both PU and SU behaviors on the availability of white space channels for CA. Unlike the widely used ON-OFF process that is oversimplified, this model is general to capture a wide range of user behaviors.
- We derive the delay costs for performing CA under this model. User demands in both the frequency and time domains are considered to evaluate the costs for making negotiation and renewing transmission.
- We derive a CA strategy that minimizes the expected cumulative delay for transmitting certain data.
- Our numerical and simulation results based on real data of primary user activity show that user demands on aggregated bandwidth and service duration should be carefully chosen.

The remainder of this chapter is organized as follows. Related work is discussed in section 3.2. System model is introduced in section 3.3. The delay costs for utilizing CA are derived in section 3.4. The optimal use of CA is proposed in section 3.5. Numerical and simulation results are presented in section 3.6. The chapter is summarized in section 3.7.

3.2 Related Work

Recently, the benefit of enabling CA in capacity expansion has been recognized by some existing cognitive MAC protocols [36, 37, 39, 40, 41, 42]. However, none of them has considered the following two facts at the same time. First, CA is only beneficial under certain patterns of PU activity. Second, there are limitations on radio capability to perform CA over a large spectrum range. Thus, these work cannot accurately capture the constraints and costs for implementing CA under the influence of PU activity in both the frequency and time domains. As a result, the possible failures of SU service caused by PU activity are not correctly reflected in [36, 37, 39, 40, 41], while the limited radio capability for performing CA is not reflected in [36, 39, 40, 41, 42]. Besides, all the work above neglect to analyze the appropriate use of CA in the time domain that also significantly affects the efficiency of CA.

Most of the existing channel usage models assume that channel idle or service duration is exponentially distributed in either frequency domain [42, 43, 44] or time domain (i.e., the ON-OFF process) [36, 37, 39, 40, 41]. Unfortunately, recent measurement study has shown that the distribution of call duration in cellular networks is not exponential and is hard to fit into any existing closed-form models [9]. In fact, modeling of PU behavior is still a very challenging job. We conjecture that the distribution of PU service duration in other types of networks may not be exponential either. Furthermore, the existing channel usage models neglect the impact of secondary user activity that is needed for coexistence analysis.

3.3 System Model

In this section, we propose a channel usage model based on general assumptions on user behaviors. Prior to this, some basic assumptions are introduced first.

3.3.1 Channel Aggregation Establishment

A secondary network is assumed to be overlaid upon a general primary network. Both the networks operate in the same licensed spectrum $\mathcal{F} = \{f_1, \cdots, f_K\}$, which consists of K channels in total. Each SU with a b-channel *bandwidth demand* can assemble b channels at a time so as to form an aggregated channel $\mathcal{A}^{(b)} = \{f_{1'}, \cdots, f_{b'}\} \subset \mathcal{F}$ for data communication. In order to limit radio complexity, let $b \leq B$ and set $B = 3$. Note that it is not hard to extend our model for the cases with $B > 3$. To further make CA realizable, another constraint is required. If any two channels, say f_l and $f_{l+\delta}$, are too far apart in $\mathcal{F}$, it would be costly to aggregate them [37]. Then, a constraint Δ is defined to limit $\delta \leq \Delta$ for CA. In other words, an $\mathcal{A}^{(b)}$ can only be selected from $\mathcal{C}_{l,\Delta} = \{f_l, \cdots, f_{l+\Delta}\} \subset \mathcal{F}$, which is a set of candidate channels satisfying such constraint.

In addition to the required CR operating on data channels, each SU is assumed to be equipped with an extra scanner radio (SR) following Microsoft's KNOWS prototype [45, 46]. Regularly, SR operates on a dedicated control channel located in unlicensed spectrum to send or receive control messages. The access of control channel can be achieved through an IEEE 802.11-like MAC protocol. Also, SR is responsible for spectrum sensing to discover local white spaces. At each SU, say n, the set of white space channels $\mathcal{W}_n \subseteq \mathcal{F}$ can be maintained, which is detected and updated in n's sensing region V_n.

As in Figure 3.1, whenever a sender n tries to send data packets to its receiver n', a *negotiation* between n and n' via control channel is necessary for an agreement on $\mathcal{A}^{(b)}$. It is assumed that each SU does not have full knowledge of spectrum usage in its vicinity. Thus, multiple negotiation attempts between n and n' may be needed. Specifically, n should first sense a set of channels to find common white space channels in both V_n and $V_{n'}$. Here, n chooses a $\mathcal{C}_{l,\Delta}$ to sense and checks whether $|\mathcal{T}_{l,\Delta}| \geq b$, where $\mathcal{T}_{l,\Delta} = \mathcal{C}_{l,\Delta} \cap \mathcal{W}_n$. If it is true, n initiates a *handshake* with n' to see whether $|\mathcal{P}_{l,\Delta}| \geq b$, where $\mathcal{P}_{l,\Delta} = \mathcal{T}_{l,\Delta} \cap \mathcal{W}_{n'}$. If it is still true, n' selects an $\mathcal{A}^{(b)} \subseteq \mathcal{P}_{l,\Delta}$ and replies to n. Then, a *transmission* $\mathcal{S}^{(b)} = \{s_{1'}, \cdots, s_{b'}\}$ is initiated, which includes b parallel subflows with a d-slot *duration demand*. Otherwise, a

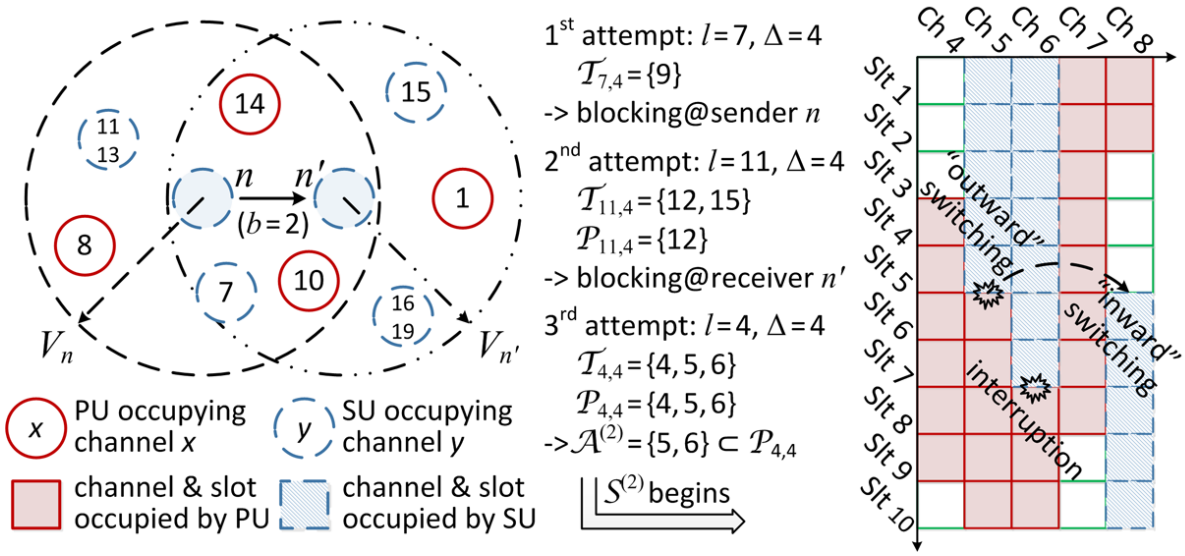

Figure 3.1: An example of channel aggregation establishment between n and n'.

blocking incident occurs due to either $|\mathcal{T}_{l,\Delta}| < b$ or $|\mathcal{P}_{l,\Delta}| < b$. The pair of n and n' have to keep checking new $\mathcal{C}_{l',\Delta}$ until $|\mathcal{P}_{l',\Delta}| \geq b$.

However, a successful negotiation does not mean a reliable transmission in the end, owing to the low priority of SU service. To overcome the service failure caused by returning PUs, spectrum switching is employed. Specifically, whenever a PU arrival to any $f_k \in \mathcal{A}^{(b)}$ is detected, the pair of n and n' needs to vacate the preempted f_k immediately and then tries to renew the corresponding s_k on a backup channel $f_{k'} \in \mathcal{B}_{l,\Delta}$, where $\mathcal{B}_{l,\Delta} = \mathcal{P}_{l,\Delta} \backslash \mathcal{A}^{(b)}$. For ease of presentation, such spectrum switching is divided into two steps: "outward" switching from f_k and "inward" switching to $f_{k'}$. If $|\mathcal{B}_{l,\Delta}| = 0$, an *interruption* incident occurs to the expelled s_k. But all the other on-going ones in $\mathcal{S}^{(b)}$ may not be affected as long as the independence of these parallel subflows is guaranteed [42].

3.3.2 Channel Usage Model

In a certain SU n's vicinity, any $f_c \in \mathcal{F}$ may be occupied by an active PU or SU service for a period of time. As in Figure 3.2, the average channel occupancy on such f_c can be

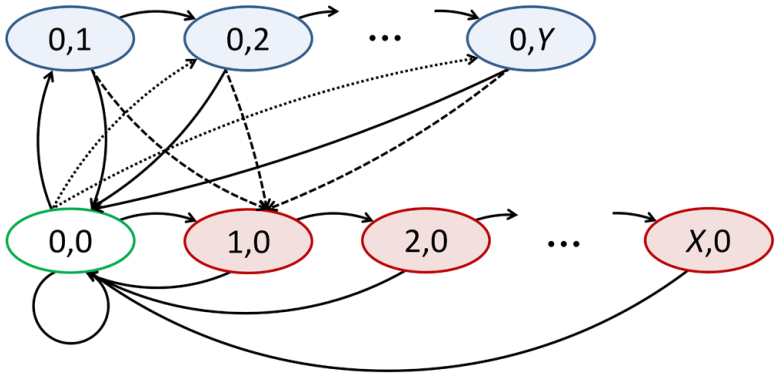

Figure 3.2: A channel usage model.

modeled as a Markov chain, in which channel state transits on a slot basis with a τ-second slot duration. Three groups of channel states are defined as follows: 1) *idle state* (0,0), in which f_c can be a white space; 2) *PU service states* $(x,0)$'s, $x \in \{1, \cdots, X\}$, in which f_c has been occupied by a PU service for x slots; 3) *SU service states* $(0,y)$'s, $y \in \{1, \cdots, Y\}$, in which f_c has been occupied by a SU subflow for y slots. Both X and Y are large enough such that $\Pr[x > X]$ and $\Pr[y > Y]$ are negligible. If there are multiple PU or SU services sharing f_c, the statistical data of the service with maximum duration can be applied.

The availability of white spaces is characterized by the steady-state probabilities of channel states, denoted by $\pi_{(\text{current state})}$'s, especially $\pi_{(0,0)}$ for idle state. To derive them, the transition probabilities, denoted by $\omega^{(\text{next state})}_{(\text{current state})}$'s, are obtained as follows.

First, state transitions from (0,0) and $(0,y)$'s to (1,0) represent a PU arrival to f_c. Each transition from $(0,y)$ to (1,0) also indicates an outward switching from f_c. The transition probabilities are actually equal to the PU arrival probability, denoted by λ_α, which further depends on the PU arrival process with average arrival rate α_n learnt in V_n. Then, we have

$$\omega^{(1,0)}_{(0,0)} = \omega^{(1,0)}_{(0,y)} = \lambda_\alpha, \ y \in \{1, \cdots, Y\}. \tag{3.1}$$

Next, state transitions from $(x,0)$'s to (0,0) and that among $(x,0)$'s are defined by the distribution of PU service duration on f_c. Note that any closed-form distribution function is not necessarily required here. Instead, one can directly input the statistical distribution of service

duration collected from a real network to determine the following transition probabilities

$$\begin{cases} \omega_{(x,0)}^{(0,0)} = \Pr\left[(x-1)\tau < S_{pu} \leq x\tau\right], \ x \in \{1, \cdots, X\}; \\ \omega_{(x,0)}^{(x+1,0)} = 1 - \omega_{(x,0)}^{(0,0)}, \ x \in \{1, \cdots, X-1\}, \end{cases} \tag{3.2}$$

where S_{pu} denotes the random variable of PU service duration.

In a similar way, state transitions from (0,y)'s to (0,0) and that among (0,y)'s are defined by the distribution of SU service duration on f_c, which can also be general. Hence, we have

$$\begin{cases} \omega_{(0,y)}^{(0,0)} = \left(1 - \omega_{(0,y)}^{(1,0)}\right) \Pr\left[(y-1)\tau < S_{su} \leq y\tau\right], \ y \in \{1, \cdots, Y\}; \\ \omega_{(0,y)}^{(0,y+1)} = 1 - \omega_{(0,y)}^{(1,0)} - \omega_{(0,y)}^{(0,0)}, \ y \in \{1, \cdots, Y-1\}, \end{cases} \tag{3.3}$$

where S_{su} denotes the random variable of SU service duration.

Further, state transition from (0,0) to (0,1) can be triggered by either a SU arrival or an inward switching to f_c. For the former, the SU arrival probability, denoted by λ_β, is determined by the SU arrival process with average arrival rate β_n learnt in V_n. As for the latter, an inward switching may trigger state transitions from (0,0) to not only (0,1) but also other (0,y)'s. Each transition from (0,0) to (0,y) indicates the case that a $s_{c'}$ switches to f_c when it has last y-1 slots on $f_{c'}$ and should be renewed on f_c starting from the y^{th} slot.

For more details, we first derive the probability that a certain subflow has successfully switched into f_c, denoted by γ. Due to the restriction placed by Δ, only the $2{\cdot}\Delta$ channels excluding f_c in $\mathcal{C}_{c-\Delta,2\Delta}$ qualify for an inward switching to f_c. If there are u preempted channels and v idle channels out of such $2{\cdot}\Delta$ channels, γ is equivalent to the probability that one of the u expelled subflows successfully chooses f_c out of the total v+1 idle channels for an inward switching. Here, the worst case is analyzed, in which any incoming subflow

neglects the idle channels that are not included in $\mathcal{C}_{c-\Delta,2\Delta}$. Then, we have

$$\begin{aligned}\gamma = \sum_{u=1}^{2\Delta}\sum_{v=0}^{2\Delta-u}\frac{(2\Delta)!}{u!v!(2\Delta-u-v)!}\left[\lambda_\alpha\left(\sum_{y=1}^{Y}\pi_{(0,y)}\right)\right]^u\left(\pi_{(0,0)}\right)^v \\ \cdot\left[1-\lambda_\alpha\left(\sum_{y=1}^{Y}\pi_{(0,y)}\right)-\pi_{(0,0)}\right]^{2\Delta-u-v}\min\left\{\tfrac{u}{v+1},1\right\}.\end{aligned} \tag{3.4}$$

Given that a subflow, say $s_{c'}$, has switched into f_c, we further derive $\xi^{(0,y)}$, the probability that $s_{c'}$ has finished y-1 slots on $f_{c'}$. For a Y-slot subflow, if an outward switching is assumed to occur in each slot with the same probability, we have

$$\xi^{(0,y)} = \sum_{z=y}^{Y}\Pr\left[(z-1)\tau < S_{su} \le z\tau\right]\frac{1}{z},\ y\in\{1,\cdots,Y\}. \tag{3.5}$$

Using (3.4) and (3.5), we have the following transition probabilities

$$\begin{cases}\omega_{(0,0)}^{(0,1)} = \left(1-\omega_{(0,0)}^{(1,0)}\right)\left(\lambda_\beta+\gamma\xi^{(0,1)}\right); \\ \omega_{(0,0)}^{(0,y)} = \left(1-\omega_{(0,0)}^{(1,0)}\right)\gamma\xi^{(0,y)},\ y\in\{2,\cdots,Y\}.\end{cases} \tag{3.6}$$

Finally, we are able to have the transition probability from (0,0) to itself

$$\omega_{(0,0)}^{(0,0)} = 1-\omega_{(0,0)}^{(1,0)}-\sum_{y=1}^{Y}\omega_{(0,0)}^{(0,y)}. \tag{3.7}$$

3.4 Costs for Channel Aggregation

Both the negotiation and transmission operations between a sender n and its receiver n' involve some kinds of service failures and some levels of delay costs. In this section, we aim to investigate the corresponding service failure probabilities and model the delay costs for performing CA under the influence of PU activity.

3.4.1 Negotiation Costs

The efficiency of negotiation between n and n' is restricted by the availability of white spaces in both V_n and $V_{n'}$ all the time. In general, the delay costs for making a successful negotiation include: i) *sensing delay* $T_{ss}^{(b)}$, which is the time required for sensing channels at both sides of n and n'; ii) *handshake delay* $T_{hs}^{(b)}$, which is the time required for accessing control channel and making handshakes on it back and forth.

Intuitively, such negotiation costs are proportional to b, i.e., the number of channels that are demanded, and r, which denotes the number of blocking incidents caused by $|\mathcal{P}_{l,\Delta}| < b$ during the entire negotiation. The blocking probability is defined as

$$\theta^{(b)} = \sum_{v=0}^{b-1} \binom{\Delta+1}{v} \left(\pi_{(0,0)}\right)^{v} \left(1-\pi_{(0,0)}\right)^{(\Delta+1)-v}, \tag{3.8}$$

in which $\pi_{(0,0)}$ is computed using $\alpha_{n,n'} = c_{n,n'} \cdot \alpha_n$ and $\beta_{n,n'} = c_{n,n'} \cdot \beta_n$ in $V_n \cup V_{n'}$, where $c_{n,n'}$ denotes a correlation factor. The entire negotiation should be given up as soon as r reaches a threshold $\hat{N}_{bl}$. Define the negotiation failure probability as

$$\varepsilon^{(b)} = 1 - \sum_{r=0}^{\hat{N}_{bl}} \left(\theta^{(b)}\right)^{r} \left(1-\theta^{(b)}\right). \tag{3.9}$$

For a successful negotiation, the expected value of r is

$$N_{bl}^{(b)} = \sum_{r=0}^{\hat{N}_{bl}} \left(\theta^{(b)}\right)^{r} \left(1-\theta^{(b)}\right) r. \tag{3.10}$$

Because n initiates a handshake only when $|\mathcal{T}_{l,\Delta}| \geq b$, we also need to obtain $\tilde{\theta}^{(b)}$, the blocking probability due to $|\mathcal{T}_{l,\Delta}| < b$. As defined in (3.8), $\tilde{\theta}^{(b)}$ is actually as same as $\theta^{(b)}$, but $\pi_{(0,0)}$ is computed using α_n and β_n in V_n here instead. Then after we are able to get $\tilde{\theta}^{(b)}$, for a

successful negotiation, the expected number of handshake attempts is

$$N_{hs}^{(b)} = N_{bl}^{(b)} \left(1 - \frac{\tilde{\theta}^{(b)}}{\theta^{(b)}}\right) + 1, \tag{3.11}$$

where $1 - \tilde{\theta}^{(b)}/\theta^{(b)}$ denotes the probability that $|\mathcal{T}_{l,\Delta}| \geq b$ but $|\mathcal{P}_{l,\Delta}| < b$. In order to get $T_{ss}^{(b)}$, sequential sensing [39, 47] is assumed to use. Specifically, whenever a new $\mathcal{C}_{l,\Delta}$ is chosen, n needs to sense the channels in it to keep $\mathcal{T}_{l,\Delta}$ fresh, while n' senses the channels in $\mathcal{T}_{l,\Delta}$ to complete the entire negotiation. We are able to have

$$T_{ss}^{(b)} = N_{hs}^{(b)} \left[\sum_{v=b}^{\Delta+1} \binom{\Delta+1}{v} \left(\pi_{(0,0)}\right)^{v} \left(1 - \pi_{(0,0)}\right)^{(\Delta+1)-v} v\right] \tau_{ss} + \left(N_{bl}^{(b)} + 1\right) (\Delta + 1)\, \tau_{ss}, \tag{3.12}$$

where τ_{ss} denotes the average time for sensing one channel, and $\pi_{(0,0)}$ is computed using α_n and β_n. Then, we are also able to have

$$T_{hs}^{(b)} = N_{hs}^{(b)} \left(\tau_{ma} + \tau_{rt}\right), \tag{3.13}$$

where τ_{ma} denotes the average time for accessing control channel, which would be given by classic analytical models [48], and τ_{rt} denotes the round-trip time for one handshake.

3.4.2 Transmission Costs

The success of transmission is not guaranteed due to the occurrence of interruption incidents. In general, the delay costs for completing a successful transmission include: i) *switching delay* $T_{sw}^{(b,d)}$, which is the time required for vacating the preempted channels and renewing the corresponding subflows; ii) *transmission delay* T_{tx}, related to data size and data rate.

Typically, such transmission costs depend not only on b but also on d, i.e., the number of slots demanded for service duration after each successful negotiation. Sometimes, dividing a large size of data into smaller segments and transmitting them separately with a shorter

duration can be a better choice. When a PU arrival to $\mathcal{A}^{(b)}$ is detected, the switching subflow in $\mathcal{S}^{(b)}$ can be interrupted if it finds $|\mathcal{B}_{l,\Delta}| = 0$. It is also possible that more than one subflow in $\mathcal{S}^{(b)}$ is interrupted in the same slot. Hence, to study such bandwidth reduction of $\mathcal{A}^{(b)}$ in each slot, we define a one-step interruption probability matrix

$$\mathbf{\Phi} = \begin{pmatrix} \varphi_{0,0} & & & \\ \varphi_{1,0} & \varphi_{1,1} & & \\ \varphi_{2,0} & \varphi_{2,1} & \varphi_{2,2} & \\ \varphi_{3,0} & \varphi_{3,1} & \varphi_{3,2} & \varphi_{3,3} \end{pmatrix},$$

in which each $\varphi_{i,j}$ denotes the probability that $i-j$ subflows in $\mathcal{S}^{(i)}$ are interrupted in one slot. Specifically, we have

$$\begin{cases} \varphi_{i,j} = \sum_{u=i-j}^{i} \binom{i}{u} (\lambda_\alpha)^u (1-\lambda_\alpha)^{i-u} \binom{(\Delta+1)-i}{u-i+j} \left(\pi_{(0,0)}\right)^{u-i+j} \\ \qquad\qquad \cdot \left(1-\pi_{(0,0)}\right)^{(\Delta+1)-u-j}, \ i \in \{1,\cdots,3\}, \ j \in \{0,\cdots,i-1\}; \\ \varphi_{i,i} = 1 - \sum_{j=0}^{i-1} \varphi_{i,j}, \ i \in \{1,\cdots,3\}. \end{cases} \quad (3.14)$$

Further for a d-slot transmission, the d-step interruption probability matrix $\mathbf{\Phi}^d = \mathbf{\Phi}^{d-1}\mathbf{\Phi}$ is used instead, in which each $\varphi_{i,j}^{(d)}$ defines the corresponding bandwidth reduction of $\mathcal{A}^{(i)}$ within d slots. Besides, $\varphi_{b,0}^{(d)}$ is referred to as the transmission failure probability. Here, we mainly derive $T_{sw}^{(b,d)}$ for a successful d-slot subflow, since the subflows in $\mathcal{S}^{(b)}$ are parallel. Given that there has been no interruption, let $\chi^{(b)}$ be the probability that a switching operation successfully occurs in one slot. We are able to have

$$\chi^{(b)} = \frac{\lambda_\alpha \left[1 - \left(1-\pi_{(0,0)}\right)^{(\Delta+1)-b}\right]}{1 - \lambda_\alpha \left(1-\pi_{(0,0)}\right)^{(\Delta+1)-b}}, \quad (3.15)$$

in which $\pi_{(0,0)}$ is computed using $\alpha_{n,n'}$ and $\beta_{n,n'}$. Note that b in (3.15) may be reduced by interruption incidents. Here, the worst case is studied by fixing b to its initial value. Then,

within d slots, the expected number of switching operations is

$$N_{sw}^{(b,d)} = \sum_{z=0}^{d} \binom{d}{z} \left(\chi^{(b)}\right)^{z} \left(1-\chi^{(b)}\right)^{d-z} z. \tag{3.16}$$

Accordingly, for a successful d-slot subflow, we have

$$T_{sw}^{(b,d)} = N_{sw}^{(b,d)} \tau_{sw}, \tag{3.17}$$

where τ_{sw} is the time required for one switching operation. The sensing time for locating backup channels can be negligible due to the simultaneous operations of CR and SR.

3.4.3 Cumulative Delay

The *cumulative delay* $T_{cm}^{(b,d)}$ for transmitting the data with a M-bit size is defined as the objective for CA. The average data that can be successfully transmitted by each attempt is

$$\tilde{M}^{(b,d)} = \left(1-\varepsilon^{(b)}\right) \left(\sum_{j=0}^{b} \varphi_{b,j}^{(d)} Rjd\tau \right), \tag{3.18}$$

where R is the bit rate on one channel. At each attempt, consider: 1) negotiation fails; 2) negotiation succeeds but transmission fails; 3) both negotiation and transmission succeed.

$$\begin{aligned} T_{cm}^{(b,d)} = \tfrac{M}{\tilde{M}^{(b,d)}} \Bigg\{ & \varepsilon^{(b)} \left(\hat{T}_{ss}^{(b)} + \hat{T}_{hs}^{(b)} \right) + \left(1-\varepsilon^{(b)}\right) \Bigg[T_{ss}^{(b)} + T_{hs}^{(b)} \\ & + \varphi_{b,0}^{(d)} \left(T_{sw}^{(b,\frac{d}{2})} + \tfrac{d}{2}\tau \right) + \left(1-\varphi_{b,0}^{(d)}\right) \left(T_{sw}^{(b,d)} + d\tau \right) \Bigg] \Bigg\}, \end{aligned} \tag{3.19}$$

where $\hat{T}_{ss}^{(b)}$ and $\hat{T}_{hs}^{(b)}$ are computed by replacing $N_{bl}^{(b)}$ with $\hat{N}_{bl}$ in (3.12) and (3.13), respectively, and the expected duration of failed service related to $\varphi_{b,0}^{(d)}$ is assumed to be $d/2$.

3.5 Channel Aggregation Strategy

As shown above, the delay costs for implementing CA are closely related to the values of b and d, i.e., the user demands on aggregated bandwidth and service duration. On the one hand, the choice of b should consider the trade-off between channel capacity and blocking (interruption) probability during a negotiation (transmission). More white spaces are needed to meet a higher requirement of b. On the other hand, the choice of d should consider the trade-off between negotiation overhead and interruption probability during a transmission. With certain data to transmit, one can choose to divide the entire data into d-slot segments. A larger value of d results in fewer data segments and thus fewer negotiation operations. However, there may be more spectrum resources wasted for nothing due to higher interruption probability. Therefore, in this section, we need to find the optimal combination of b and d to achieve the efficiency of CA with confidence.

Now, a CA strategy (b, d) is defined as the combination of both bandwidth and duration demands. Then, our objective is to find the optimal (b^*, d^*) that minimizes $T_{cm}^{(b,d)}$:

$$(b^*, d^*) = \arg\min_{(b,d)\in\mathcal{G}} T_{cm}^{(b,d)}. \tag{3.20}$$

It is not hard to find (b^*, d^*) by searching $\mathcal{G}$, which denotes the finite set that contains all possible (b, d)'s. Note that in CR networks, both PU and SU behaviors that affect the availability of white spaces are stochastic. Hence, the optimal CA strategy defined in (3.20) is actually optimal in the sense of average performance.

For practical purpose, each n needs to manage a database that records the recent history of sensing results in V_n besides the real-time updated $\mathcal{W}_n$. Also, each n needs to coordinate with its every neighbor n' to learn $c_{n,n'}$ based on their local observations. In this way, n can dynamically adjust its inputs to our model and adapt (b^*, d^*) accordingly. For better real-time performance, the optimal solution can be reached by searching only a subset of $\mathcal{G}$ following the previous decisions or by searching a decision table that is built offline.

3.6 Performance Evaluation

In order to validate the effectiveness of our model and the above optimal CA strategy, both numerical analysis and discrete-event simulation are conducted in this section.

Here, we focus on a pair of sender and receiver, n and n', in a distributed CR network to evaluate the efficiency of CA between them. For PU behavior in both V_n and $V_{n'}$, Poisson PU arrival process is assumed. The distribution of PU service duration is set according to the statistical distribution of call duration collected from a real cellular network [9]. As for SU behavior, Poisson SU arrival process and random SU service duration are assumed. Note that the choice of service duration for the pair of n and n' is a part of their optimal decision, but we fix the patterns of SU activity in the background.

We are interested in the impact of PU activity and CA strategy on the efficiency of negotiation and transmission between n and n'. The constant parameters are set as follows: $K = 50$; $\beta_n = 0.02$ user/s; $c_{n,n'} = 1.5$; $E[Y] = 3$ s; $\tau = 10$ ms; $\tau_{ss} = 10$ ms [46, 47]; $\tau_{rt} = 200$ ms; $\tau_{sw} = 600$ ms [39]; $\hat{N}_{bl} = 5$; $M = 50$ Mb; $R = 5$ Mb/s. The others are viewed as variables.

As in (3.9), $\varepsilon^{(b)}$ defined for negotiation failure is connected to $\theta^{(b)}$ directly that characterizes the occurrence of blocking incidents caused by $|\mathcal{P}_{l,\Delta}| < b$. The impact of α_n and Δ on $\varepsilon^{(b)}$ with fixed $d{\cdot}\tau = 3$ s is plotted in Figure 3.3. Generally, the numerical and simulation results match well with each other under the same settings. As what we have expected, the higher demand on b the higher $\varepsilon^{(b)}$ has to be experienced. And $\varepsilon^{(b)}$ climbs up to a high level as the availability of white spaces drops. In addition, a relaxation of the hardware limitation on Δ offers more candidate channels and thus lowers down $\varepsilon^{(b)}$.

Moreover, any subflow between n and n' may be terminated halfway under the occurrence of interruption incidents. The detailed bandwidth reduction caused by interruption is characterized by $\mathbf{\Phi}^d$ based on (3.14). Due to limited space, only the impact of d on $\varphi_{b,0}^{(d)}$ defined for transmission failure is shown at this point. As in Figure 3.4 with fixed $\alpha_n = 0.1$ user/s, the numerical and simulation results match well. Clearly, the higher demand on d the higher $\varphi_{b,0}^{(d)}$

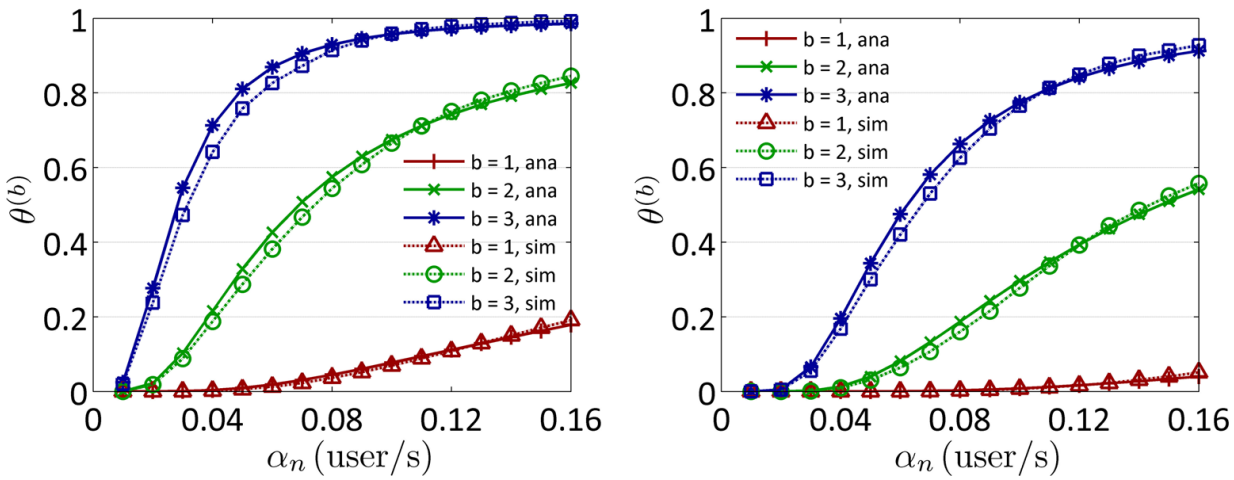

Figure 3.3: Impact of PU arrival rate on negotiation failure probability: a) $\Delta = 10$ (left); b) $\Delta = 20$ (right).

has to be experienced. And Δ also has its impact. However, unlike $\varepsilon^{(b)}$, a higher demand on b achieves even better $\varphi_{b,0}^{(d)}$, since the transmission consisting of more subflows would tolerant more interruption incidents. The impact of α_n on $\varphi_{b,0}^{(d)}$ is similar to that on $\varepsilon^{(b)}$.

For the transmission of M-bit data, the related cumulative delay $T_{cm}^{(b,d)}$ has been chosen as our objective function as in (3.19). In Figure 3.5 and Figure 3.6, the impact of α_n and (b, d) on it is plotted, respectively. It can be seen that $T_{cm}^{(b,d)}$ rises rapidly with the increase of α_n under the influence of $\varepsilon^{(b)}$ and $\varphi_{b,0}^{(d)}$, especially when $b = 3$. The greater radio capability on Δ also lowers down $T_{cm}^{(b,d)}$. To achieve the lowest $T_{cm}^{(b,d)}$, the optimal CA strategy (b^*, d^*) is evaluated under different settings. In Figure 3.6, the marked point that represents the optimal decision varies significantly with the availability of white spaces. Obviously, when there are plenty of white spaces as in Figure 3.6.a, the high demands on both b and d can be the optimal solution to achieve the highest utilization of licensed spectrum. Note that the range of d differs for different values of b for transmitting the same size of data. However, if there are few white spaces as in Figure 3.6.c, the demands on both b and d should be low to avoid the huge costs for making negotiation and renewing transmission.

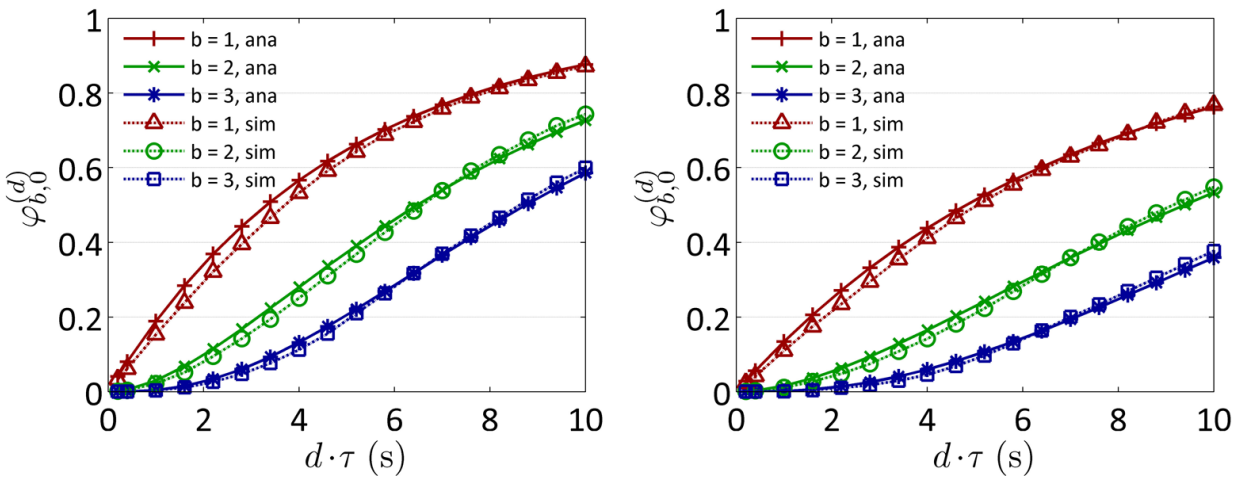

Figure 3.4: Impact of duration demand on transmission failure probability: a) $\Delta = 10$ (left); b) $\Delta = 20$ (right).

3.7 Chapter Summary

In this chapter, we have studied the efficiency of CA in consideration of dynamic spectrum availability and limited radio capability. A channel usage model based on general assumptions on user behaviors has been proposed to investigate the negotiation and transmission costs for utilizing CA under the influence of PU activity. We have found that user demands on both aggregated bandwidth and service duration affect the delay performance a lot. Hence, an optimal CA strategy has been derived and validated to achieve the lowest cumulative delay for transmitting certain data with confidence.

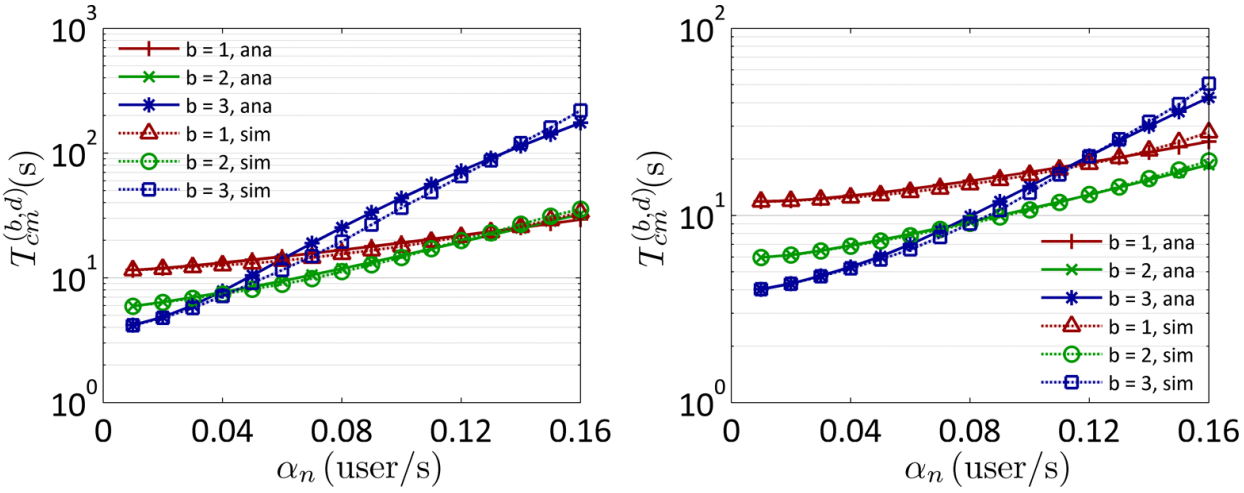

Figure 3.5: Impact of PU arrival rate on cumulative delay (in log scale): a) $\Delta = 10$ (left); b) $\Delta = 20$ (right).

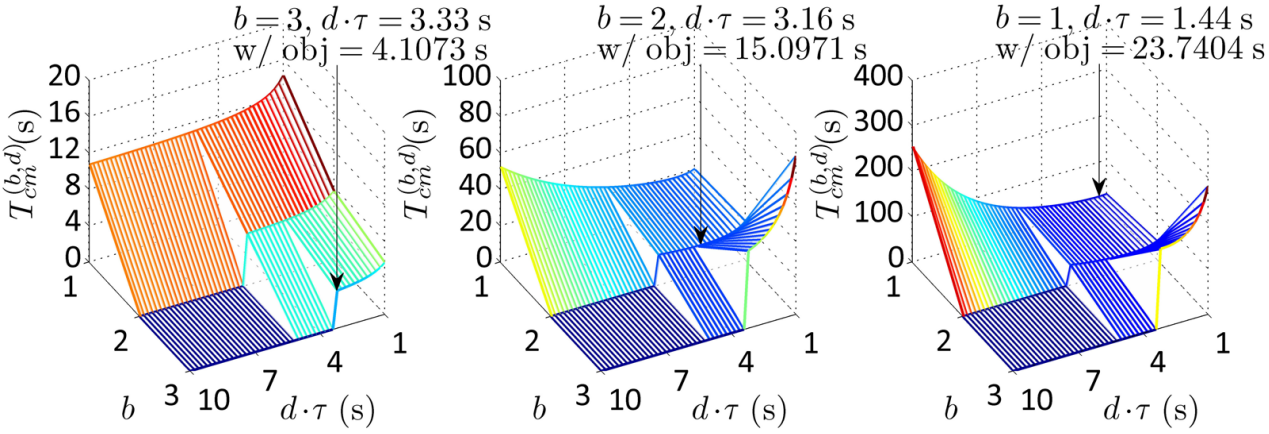

Figure 3.6: Impact of bandwidth-duration demands on cumulative delay: a) $\alpha_n = 0.01$ user/s (left); b) $\alpha_n = 0.1$ user/s (middle); c) $\alpha_n = 0.16$ user/s (right).

Chapter 4

Uplink Soft Frequency Reuse for Self-Coexistence of CR Networks

In this chapter, we focus on the second scenario to study the coexistence of homogeneous CR networks, termed as self-coexistence, when co-channel networks do not rely on inter-network coordination [13, 14].

4.1 Challenges and Contributions

The opportunistic use of licensed spectrum by secondary users equipped with CRs and the resulting coexistence between primary users and secondary users have given rise to a number of challenging technical problems. Thus far, the problem of incumbent protection has received the most attention, and has been addressed by the use of incumbent geolocation databases augmented with spectrum sensing techniques. However, in contrast, the problem of network self-coexistence is not yet well understood. Some recent efforts have been made to support the coexistence of CR networks, such as IEEE 802.19.1 in TVWSs [15] and small cell coexistence in 3.5 GHz band [18, 19]. However, a dedicated centralized coexistence system

is not readily available for common cases where distributed CR networks share white space spectrum (or unlicensed spectrum) with little or no direct coordination.

In the worst case where there are an insufficient number of channels to accommodate the co-located CR networks operating in the same spectrum, e.g., in dense urban environments, *co-channel spectrum sharing* among the coexisting networks is a necessity. However, standardized self-coexistence mechanisms in most wireless standards for CR networks, e.g., IEEE 802.16h [11] and IEEE 802.22 [12], are usually inefficient. For example, IEEE 802.22 only allows coexisting network cells to take turns to access the shared channel in time division without causing inter-cell interference. In fact, however, it is possible that some of the user terminals in different cells share the same channel simultaneously with acceptable levels of mutual interference. In the popular wireless standards, including IEEE 802.16h and IEEE 802.22, orthogonal frequency division multiple access (OFDMA) is widely used. In the OFDMA downlink, a soft frequency reuse (SFR) technique [49, 50, 51, 52, 53] has been proposed for co-channel self-coexistence. However, the problem of achieving efficient co-channel spectrum sharing in the *uplink* has yet to be addressed adequately.

In this chapter, we address the self-coexistence problem in the OFDMA uplink by proposing an uplink SFR (USFR) technique. However, designing such a technique is challenging due to the following issues. First, USFR has to be spectrum-efficient in a dynamic interference environment. Unlike the downlink case where inter-cell interference caused by downlink signals is relatively static, uplink interference caused by user-generated signals is much more dynamic and unpredictable. The interference environment that USFR has to consider becomes even more complex when one considers mobile/portable wireless devices, e.g., IEEE 802.16e and upcoming IEEE 802.22a [54] devices. Second, USFR has to be carried out in a distributed manner for CR networks. For most licensed networks, e.g., cellular networks, inter-cell spectrum sharing and interference issues are centrally controlled and managed by a single operator. In contrast, there is no central entity that can address those issues when multiple unlicensed CR networks coexist in the same spectrum. Third, USFR has to be globally power-efficient and locally fair when mobile/portable terminals are considered. The

battery-powered user terminals are usually energy-constrained. Furthermore, it is unfair that user terminals with equal priority consume largely different amounts of transmit power for the same signal-to-interference-plus-noise ratio (SINR), depending on whether they are close to or far away from their base station (BS).

To address the above challenges, the contributions of this chapter are summarized as follows.

- We formulate uplink resource allocation (URA) in each network cell as an optimization problem, which is further decoupled into two subproblems: subchannel allocation (SCA) and transmit power control (TPC).
- For TPC, we provide a unique optimal solution. We frame multi-cell TPC as a non-cooperative game, and prove that the Nash equilibrium can be established without inter-cell coordination. For SCA, which requires perfect global knowledge, we present a low-complexity heuristic. We also frame multi-cell SCA as a non-cooperative game.
- We integrate the SCA and TPC games as the URA game, and design a local heuristic algorithm that is a distributed two-level game-theoretic approach. We also prove the stability of the proposed algorithm.
- Our simulation results in both multi-operator and single-operator coexistence scenarios show that the proposed technique is mostly near-optimal and improves self-coexistence in spectrum utilization, power consumption and efficiency, and intra-cell fairness.

The remainder of this chapter is organized as follows. Related work is discussed in section 4.2. The coexistence framework and related problem formulation are introduced in section 4.3. The decoupled TPC and SCA subproblems are studied in sections 4.4 and 4.5, respectively. A two-level game-theoretic approach is proposed in section 4.6. Simulation results are presented in section 4.7. The chapter is summarized in section 4.8.

4.2 Related Work

There is a considerable amount of existing work on the self-coexistence of OFDMA systems. In the downlink case, several distributed SFR-based mechanisms are proposed in [51, 52, 53]. Some ideas can be borrowed from these distributed mechanisms in the downlink. For example, the intuition behind the self-organizing SFR proposed in [51] is that each cell minimizing its total power usage will tend to assign its cell-edge users into its "good" subchannels (i.e., allocate higher power levels to its "good" subchannels), thus making these subchannels "bad" for other coexisting cells. The coexisting cells will try to avoid assigning their cell-edge users to those locally "bad" subchannels, thus making the subchannels even "better" for the cell who thinks they are "good". In our design of USFR, the cell-edge users in neighboring cells can also be separated by taking different subchannels, since the cell-edge users in each cell have higher priority to take "good" subchannels and thus avoid using "bad" subchannels.

However, there is very limited work applying SFR in the uplink. In [55], an adaptive SFR-based resource allocation mechanism is proposed to study the uplink inter-cell interference coordination (ICIC) problem. The time-frequency resource blocks are adaptively reserved for either cell-center users or cell-edge users in each cell. If a cell is in short of cell-edge resource blocks, it can send out a request to "borrow" cell-edge resource blocks from its neighboring cells. Hence, inter-cell coordination is necessary to achieve co-channel spectrum sharing. However, we focus on the worst case where inter-cell coordination is not available among the coexisting cells managed by different operators. In [56], a semi-autonomous SFR-based algorithm is proposed to maximize uplink cell throughput, similar to that in [51, 52]. However, neither power consumption and efficiency nor intra-cell fairness is well considered, which should be more important in the uplink case. In [57], several heuristic SFR-based scheduling mechanisms are roughly compared in terms of uplink outage probability. However, they do not guarantee to create spectrum-efficient and power-efficient coexistence patterns. In contrast, our approach can generate near-optimal and stable coexistence patterns in a fully distributed manner. Similar to [55], uplink ICIC problem is also studied in [58, 59],

in which inter-cell coordination supports the sharing of resource blocks between cell-edge users in a cell and cell-center users in another cell. The neighboring cells can coordinate interference limits. A resource block with a high interference threshold allows cell-edge users in other cells to take, while a resource block with a low interference threshold can only be taken by cell-center users in other cells. However, our approach is based on local interference measurements and does not rely on such inter-cell coordination.

As for developing distributed mechanisms, non-cooperative games can be defined to study the multi-cell resource allocation in both downlink [60, 61] and uplink [62, 63, 64, 65], and the self-coexistence of CR networks [66, 67]. In the uplink case, however, some types of inter-cell coordination are still needed, such as the virtual referee in [62], the centralized integer program in [63], and the correlated equilibrium in [65]. In general, no existing work aims to take advantage of SFR to jointly achieve spectrum utilization, power consumption and efficiency, and intra-cell fairness all together.

4.3 System Model

In this section, we discuss the system model that underpins the proposed USFR technique as well as related issues.

4.3.1 Uplink Soft Frequency Reuse

In this chapter, we assume a spectrum environment in a densely populated urban area, where co-channel spectrum sharing is needed to accommodate the demands of co-located CR networks operating in limited white space spectrum (or unlicensed spectrum). In particular, we focus on the problem of forced sharing of a single channel among multiple uncoordinated CR network cells managed by different network operators. Each operator manages a network consisting of multiple cells, and coordinates the cells to avoid harmful inter-cell interference.

We are especially interested in the worst case where each co-located network cell in the area of interest belongs to a different operator. In such case, global network deployment and frequency planning are not available, and co-channel spectrum sharing without inter-cell coordination is challenging. Besides the above multi-operator case, in section 4.6 for implementation variants of USFR, we further study the single-operator scenario, in which all the cells in the area of interest are centrally managed by a single operator. In each cell, user terminals, either fixed or mobile/portable, are controlled by their home BS. The commonly shared channel further consists of a number of orthogonal subchannels. Each BS is able to allocate certain subcarriers on each uplink subchannel via OFDMA. We recognize that wireless channels can have different physical-layer characteristics, such as fading, shadowing, and noise [68]. However, in this chapter, we focus on the sharing of a single channel and are particularly interested in the impact of inter-cell interference (in the category of noise) on resource allocation. Other characteristics of shared narrowband subchannels are out of the scope of this chapter. The shared channel is a white space channel (or unlicensed channel), which is locally available for all the cells in the area of interest. These co-located cells are considered as co-channel cells, in which active sessions share the same available channel. We do not address the issue of incumbent protection, as this is beyond the scope of this chapter.

The idea of SFR proposed in the downlink is illustrated in Figure 4.1.a. In each of the co-channel cells, the user terminals next to their home BS, called inner users, can fully occupy the entire common channel. But the user terminals far away from their home BS, called edge users, have to take an exclusive set of subchannels in each cell. Similarly, in Figure 4.1.b, the idea of USFR is to allow some users in a cell to share certain subchannels with some users in other cells. However, things are more complex in the uplink case. We still follow the same definitions of inner users and edge users. Furthermore, uplink edge users can be either *near* or *far* ones based on whether or not they are located close to any inter-cell overlapping areas. Hence, there can be a greater number of possible coexistence patterns, i.e., resource allocation patterns, created in the uplink by the combinations of various users.

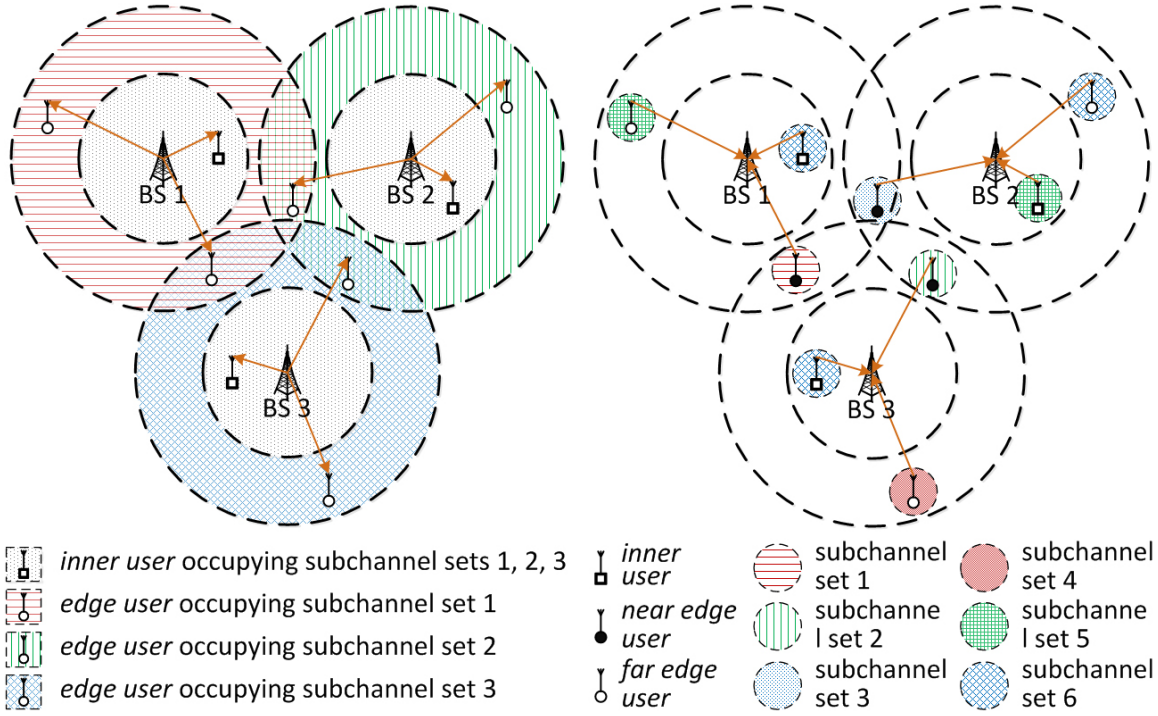

Figure 4.1: Basic ideas: a) SFR (left); b) USFR (right).

4.3.2 Uplink Resource Allocation

In each network cell, the BS conducts URA for the users under control. Local URA includes SCA and TPC for the users' uplink sessions. Whenever a new session becomes active in any cell, its home BS needs to redo URA to spare enough resource for the new session. Due to the dynamic interference environment where spectrum availability can be temporary and user terminals can be mobile/portable, we focus on a certain short time period, in which the interference environment is relatively static without the need for making prediction.

Suppose that there are total N cells coexisting on a common channel, which consists of K subchannels. In each cell n for $n \in \mathcal{N} \triangleq \{1, \cdots, N\}$, there are $M^{(n)}$ active sessions. Each user in cell n can maintain multiple sessions, and each session m_n for $m_n \in \mathcal{M}^{(n)} \triangleq \{1, \cdots, M^{(n)}\}$ can operate on multiple subchannels. We want to make the common channel accommodate all the uplink sessions that are active in the coexisting cells, but it is still possible that some sessions have to be blocked if the BS cannot find a feasible solution to the

URA problem defined below. This issue is characterized by spectrum utilization, which will be discussed in section 4.7. To avoid unnecessary intra-cell interference, each subchannel k for $k \in \mathcal{K} \triangleq \{1, \cdots, K\}$ cannot be assigned to more than one session in the same cell at the same time [55], since the sessions share the same home BS as a common receiver in the uplink. Because we focus on each short time period, the same subchannel can still be allocated to multiple sessions in different time periods when USFR is carried out periodically.

In the cell $n = a$, local URA strategy is characterized by SCA and TPC strategy matrices, which are denoted by $\mathbf{U}^{(a)} \triangleq \{U^{(a)}_{m_a,k}\}_{M^{(a)} \times K}$ and $\mathbf{P}^{(a)} \triangleq \{P^{(a)}_{m_a,k}\}_{M^{(a)} \times K}$, respectively. Each $U^{(a)}_{m_a,k}$ in $\mathcal{U}^{(a)} \triangleq \{\mathbf{U}^{(a)}\}$ denotes a binary indicator that is equal to 1 (0) when session m_a takes (does not take) subchannel k, and each $P^{(a)}_{m_a,k}$ in $\mathcal{P}^{(a)} \triangleq \{\mathbf{P}^{(a)}\}$ denotes the corresponding allocated transmit power. Each pair of $U^{(a)}_{m_a,k}$ and $P^{(a)}_{m_a,k}$ has the following relationship: $P^{(a)}_{m_a,k} > 0$ when $U^{(a)}_{m_a,k} = 1$, while $P^{(a)}_{m_a,k} = 0$ when $U^{(a)}_{m_a,k} = 0$. The singleton matrix sets $\mathcal{U}^{(a)}$ and $\mathcal{P}^{(a)}$ include the elements of matrices $\mathbf{U}^{(a)}$ and $\mathbf{P}^{(a)}$, respectively. The set of URA strategies in the other $N-1$ cells is characterized by matrix sets $\mathcal{U}^{-a} \triangleq \dot{\times}_{n \in \mathcal{N}, n \neq a} \mathcal{U}^{(n)}$ and $\mathcal{P}^{-a} \triangleq \dot{\times}_{n \in \mathcal{N}, n \neq a} \mathcal{P}^{(n)}$, where $\dot{\times}$ represents the Cartesian product [69].

Ideally, the local URA in the cell a can be formulated as a global optimization problem. The cell a's local objective in terms of power consumption, $L^{(a)}$, is defined as

$$L^{(a)} \triangleq \sum_{m_a=1}^{M^{(a)}} w^{(a)}_{m_a} \sum_{k=1}^{K} P^{(a)}_{m_a,k}, \tag{4.1}$$

where each $w^{(a)}_{m_a}$ denotes session m_a's weight or priority. In consideration of the intra-cell fairness among inner users, near edge users, and far edge users, we define $w^{(a)}_{m_a}$ as

$$w^{(a)}_{m_a} \triangleq \frac{\sum_{n=1, n \neq a}^{N} H^{(a,n)}_{m_a}}{H^{(a,a)}_{m_a}}, \tag{4.2}$$

where each $H^{(a,n)}_{m_a}$ denotes the propagation gain from session m_a to BS n. The definition of user weights is not limited to (4.2). We will discuss alternate definitions in section 4.6.

There are several constraints for the minimization of $L^{(a)}$. First, if $U^{(a)}_{m_a,k} = 1$, then $P^{(a)}_{m_a,k}$ should be lower bounded by $\hat{Q}^{(a)}_{m_a,k}$, the minimum power for meeting session m_a's SINR requirement, denoted by $\gamma^{(a)}_{m_a}$. Moreover, $P^{(a)}_{m_a,k}$ should also be upper bounded by $\bar{Q}^{(a)}_{m_a,k}$, the maximum power of session m_a on each subchannel k. But if $U^{(a)}_{m_a,k} = 0$, then $P^{(a)}_{m_a,k} = 0$. The BS a cannot make decisions for the other cells to change $\mathcal{U}^{-a}$, so we say $\mathcal{U}^{-a} \equiv \tilde{\mathcal{U}}^{-a}$, where $\tilde{\mathcal{U}}^{-a}$ is a fixed matrix set. But $\mathcal{P}^{(a)}$ and $\mathcal{P}^{-a}$ may interact with each other due to the change of inter-cell interference and corresponding power control, so both $\mathcal{P}^{(a)}$ and $\mathcal{P}^{-a}$ are considered as variables. For example, when session m_a in the cell a and session $m_{a'}$ in another cell a' share the same subchannel k, i.e., $U^{(a)}_{m_a,k} = U^{(a')}_{m_{a'},k} = 1$, increasing $P^{(a)}_{m_a,k}$ can result in an increase in $P^{(a')}_{m_{a'},k}$ to maintain session $m_{a'}$'s SINR due to the increased interference from session m_a. Hence, such set of constraints should be satisfied not only in the cell a but also in other coexisting cells n, namely

$$U^{(n)}_{m_n,k}\hat{Q}^{(n)}_{m_n,k} \leq P^{(n)}_{m_n,k} \leq U^{(n)}_{m_n,k}\bar{Q}^{(n)}_{m_n,k} \text{ for } n \in \mathcal{N}; m_n \in \mathcal{M}^{(n)}; k \in \mathcal{K}, \tag{4.3}$$

where $\hat{Q}^{(n)}_{m_n,k} \triangleq \frac{\gamma^{(n)}_{m_n}(\sum_{n'=1,n'\neq n}^{N}\sum_{m_{n'}=1}^{M^{(n')}} P^{(n')}_{m_{n'},k}H^{(n',n)}_{m_{n'}}+Z_k)}{H^{(n,n)}_{m_n}}$, in which Z_k denotes the noise power on subchannel k. At each BS n, let $I^{(n)}_k \triangleq \sum_{n'=1,n'\neq n}^{N}\sum_{m_{n'}=1}^{M^{(n')}} P^{(n')}_{m_{n'},k}H^{(n',n)}_{m_{n'}}$ denote the aggregated interference from the sessions in other cells n' on subchannel k, which depends on $\mathcal{U}^{-n}$ and $\mathcal{P}^{-n}$. Second, each session m_n's aggregated uplink capacity (per unit bandwidth) in cell n should meet its corresponding QoS requirement, denoted by $\theta^{(n)}_{m_n}$. We have

$$\sum_{k=1}^{K}\log(1+\frac{P^{(n)}_{m_n,k}H^{(n,n)}_{m_n}}{I^{(n)}_k+Z_k}) \geq \theta^{(n)}_{m_n} \text{ for } n \in \mathcal{N}; m_n \in \mathcal{M}^{(n)}. \tag{4.4}$$

Third, the BS a should not assign more than one session to any subchannel k at the same time to address the issue of intra-cell interference in the uplink. We have

$$\sum_{m_a=1}^{M^{(a)}} U^{(a)}_{m_a,k} \leq 1 \text{ for } k \in \mathcal{K}. \tag{4.5}$$

The URA problem in the network cell a towards USFR is defined as follows.

Problem 1 (URA)

$$\begin{aligned} \text{Given:} \quad & \mathcal{U}^{-a}; \\ \text{Find:} \quad & \mathcal{U}^{(a)}, \mathcal{P}^{(a)}, \mathcal{P}^{-a}; \\ \text{Minimize:} \quad & L^{(a)}; \\ \text{Subject to:} \quad & (4.3), (4.4), (4.5). \end{aligned}$$

The Problem 1 is formulated as a mixed-integer non-linear program (MINLP), which is NP-hard in general [70]. Instead of solving it directly, as in [60], we can decouple the complex URA problem in the cell a into two subproblems: TPC by adapting $\mathcal{P}^{(a)}$ and $\mathcal{P}^{-a}$ given fixed $\mathcal{U}^{(a)}$, and SCA by adapting $\mathcal{U}^{(a)}$ given known $\mathcal{P}^{(a)}$ and $\mathcal{P}^{-a}$ as functions of $\mathcal{U}^{(a)}$. Hence, the URA problem can be reduced to the SCA subproblem, which is a binary integer program (BIP). In section 4.5, we show that finding the optimal solution to the SCA subproblem requires perfect global knowledge and high computational complexity.

4.3.3 A Game-Theoretic Coexistence Framework

In view of the distributed nature of CR networks, each cell in the multi-cell system has to conduct local URA independently. Because of possible conflicts in coexisting cells' local optimal strategies, we are interested in answering the following questions: 1) Is such a non-cooperative multi-cell decision-making system stable? 2) If the system is stable, how close is the distributed stable point to the global optimal point? 3) Are there any other benefits of optimizing local objective, i.e., power consumption, in each cell, such as spectral efficiency and coexistence fairness? 4) Can the non-cooperative approach be applied to the cooperative case where the coexisting cells coordinate with each other? Therefore, we choose to use game theory to study the global performance of multi-cell URA.

The self-coexistence of CR networks can be modeled as a non-cooperative game, in which each network cell acts as a player. In the URA game, each cell solves Problem 1 independently. Minimizing $L^{(n)}$ is equivalent to optimizing cell n's utility. According to the decoupled SCA

and TPC subproblems, we adopt a two-level game-theoretic approach to enable USFR. The URA game can be decoupled into SCA and TPC games. In the two-level framework, each acting cell plays the SCA game on the first level. Given each new strategy taken by any cell in the SCA game, the Nash equilibrium is achieved in the TPC game on the second level. Hence, solving the local URA problem in any acting cell is equivalent to solving the SCA subproblem once while solving the TPC subproblem multiple times (once for each change of SCA strategy by the cell). The acting cells take turns to solve the URA problem, so the URA game is achieved by successive local searches of SCA and TPC strategies optimizing local objectives. Based on the optimal utility gain, each acting cell is able to know whether this two-level URA strategy is beneficial. As soon as nobody can find an improving strategy, a stabilized coexistence pattern is commonly agreed by all the cells in the URA game.

4.4 Game of Transmit Power Control

In this section, we provide a unique optimal solution to the TPC subproblem, and study multi-cell TPC as a non-cooperative game.

4.4.1 Transmit Power Control

The existence of binary variables in $\mathcal{U}^{(a)}$ makes Problem 1 costly to solve. If $\mathcal{U}^{(a)}$ is fixed in advance, the URA problem can be reduced to the TPC subproblem. The optimal TPC strategy set, say $\tilde{\mathcal{P}}^N$, depends on the setting of $\mathcal{U}^{(a)}$, say $\mathcal{U}^{(a)} \equiv \tilde{\mathcal{U}}^{(a)}$. In our two-level framework, $\tilde{\mathcal{U}}^{(a)}$ is a SCA strategy taken to solve the SCA subproblem. Thus, $\tilde{\mathcal{U}}^{(a)}$ and $\tilde{\mathcal{P}}^N$ together should satisfy the constraints of Problem 1. Besides (4.3), (4.4), and (4.5) for $a \in \mathcal{N}$, the setting of $\mathcal{U}^{(a)} \equiv \tilde{\mathcal{U}}^{(a)}$ also needs to satisfy the conditions (4.6) and (4.8) below to make $\tilde{\mathcal{P}}^N$ feasible to Problem 1. We have

$$U^{(n)}_{m_n,k}\tilde{Q}^{(n)}_{m_n,k} \leq U^{(n)}_{m_n,k}\bar{Q}^{(n)}_{m_n,k} \text{ for } n \in \mathcal{N}; m_n \in \mathcal{M}^{(n)}; k \in \mathcal{K}, \tag{4.6}$$

where $\tilde{Q}_{m_n,k}^{(n)} = \hat{Q}_{m_n,k}^{(n)}|_{P_{m_{n'},k}^{(n')}=\tilde{P}_{m_{n'},k}^{(n')}}$ and each $\tilde{P}_{m_n,k}^{(n)}$ is from the solution to the equations

$$P_{m_n,k}^{(n)} \equiv U_{m_n,k}^{(n)} \hat{Q}_{m_n,k}^{(n)} \text{ for } n \in \mathcal{N}; m_n \in \mathcal{M}^{(n)}; k \in \mathcal{K}. \tag{4.7}$$

The values of $\tilde{P}_{m_n,k}^{(n)}$ are determined as long as $\tilde{\mathcal{U}}^{(a)}$ is fixed, and so are the values of $\tilde{Q}_{m_n,k}^{(n)}$. Each $\tilde{Q}_{m_n,k}^{(n)}$ is the minimum possible value of corresponding $\hat{Q}_{m_n,k}^{(n)}$ under $\mathcal{U}^{(a)} \equiv \tilde{\mathcal{U}}^{(a)}$. And

$$\sum_{k=1}^{K} \log(1 + U_{m_n,k}^{(n)} \gamma_{m_n}^{(n)}) \geq \theta_{m_n}^{(n)} \text{ for } n \in \mathcal{N}; m_n \in \mathcal{M}^{(n)}. \tag{4.8}$$

Hence, the TPC subproblem in the cell a is defined as follows.

Problem 2 (TPC)

Given:	$\mathcal{U}^{(a)}, \mathcal{U}^{-a}$;
Find:	$\mathcal{P}^{(a)}, \mathcal{P}^{-a}$;
Minimize:	$L^{(a)}$;
Subject to:	(4.3), (4.4), (4.5), (4.6), (4.8).

Lemma 1 Given that the previously fixed SCA strategy set $\tilde{\mathcal{U}}^N = \tilde{\mathcal{U}}^{(a)} \dot{\times} \tilde{\mathcal{U}}^{-a}$ satisfies (4.5) for $a \in \mathcal{N}$, (4.6), and (4.8), the unique optimal solution of TPC strategy set, say $\tilde{\mathcal{P}}^N = \tilde{\mathcal{P}}^{(a)} \dot{\times} \tilde{\mathcal{P}}^{-a}$, to Problem 2 satisfies (4.7).

Proof According to (4.1) and (4.2), we know that

$$\frac{\partial L^{(a)}}{\partial P_{m_a,k}^{(a)}} = w_{m_a}^{(a)} > 0 \text{ for } m_a \in \mathcal{M}^{(a)}; k \in \mathcal{K}. \tag{4.9}$$

Obviously, in order to minimize $L^{(a)}$, each $P_{m_a,k}^{(a)}$ in $\mathcal{P}^{(a)}$ needs to be as small as possible. We can see that any inequality relationship $P_{m_a,k}^{(a)} \geq U_{m_a,k}^{(a)} \hat{Q}_{m_a,k}^{(a)}$ in (4.3) can be rewritten as

$$P_{m_a,k}^{(a)} \geq \sum_{n=1,n\neq a}^{N} \sum_{m_n=1}^{M^{(n)}} c_{m_n,m_a,k}^{(n,a)} P_{m_n,k}^{(n)} + c_{m_a,k}^{(a)} \text{ for } m_a \in \mathcal{M}^{(a)}; k \in \mathcal{K}, \tag{4.10}$$

in which all the coefficients c's are non-negative. Furthermore, for each $P^{(n)}_{m_n,k}$ on the right-hand side (RHS) of (4.10), we have

$$P^{(n)}_{m_n,k} \geq \sum_{n'=1,n'\neq n}^{N} \sum_{m_{n'}=1}^{M^{(n')}} c^{(n',n)}_{m_{n'},m_n,k} P^{(n')}_{m_{n'},k} + c^{(n)}_{m_n,k} \text{ for } n \in \mathcal{N}, n \neq a; m_n \in \mathcal{M}^{(n)}; k \in \mathcal{K}, \quad (4.11)$$

in which all the coefficients c's are non-negative. We can see that any $P^{(n)}_{m_n,k}$ in $\mathcal{P}^{(n)}$ is lower bounded by a linear combination of $P^{(n')}_{m_{n'},k}$ in $\mathcal{P}^{-n}$ with all non-negative coefficients. All the inequality relationships in (4.10) and (4.11) are in a cycle, since each $P^{(a)}_{m_a,k}$ on the left-hand side (LHS) of (4.10) appears on the RHS of (4.11) as well. Clearly, when the equalities hold for all in (4.10) and (4.11), each $P^{(n)}_{m_n,k}$ in $\mathcal{P}^N = \mathcal{P}^{(a)} \dot{\times} \mathcal{P}^{-a}$ reaches its lower bound $U^{(n)}_{m_n,k}\hat{Q}^{(n)}_{m_n,k} = U^{(n)}_{m_n,k}\tilde{Q}^{(n)}_{m_n,k}$. At the same time, the weighted sum $L^{(a)}$ is minimized by $\tilde{\mathcal{P}}^N$ that solves (4.7) without considering any other constraints. Note that (4.7) is a system of linear equations, thus $\tilde{\mathcal{P}}^N$ is unique.

We next verify the feasibility of $\tilde{\mathcal{P}}^N$, the unique solution that minimizes $L^{(a)}$. We know that (4.3) holds given (4.6) and (4.7). If (4.7) holds, the LHS's of (4.4) and (4.8) are same, so (4.4) holds given (4.7) and (4.8). And (4.5) for $a \in \mathcal{N}$ is known to hold. Hence, $\tilde{\mathcal{P}}^N$ satisfies the constraints of Problem 1, and is thus the unique optimal solution to Problem 2. □

Besides minimizing power consumption as defined by $L^{(a)}$, we are also interested in maximizing the sessions' power efficiency, i.e., uplink capacity per unit power. Again in the form of weighted sum, the cell a's local objective in terms of power efficiency, $E^{(a)}$, is defined as

$$E^{(a)} \triangleq \sum_{m_a=1}^{M^{(a)}} v^{(a)}_{m_a} \sum_{\substack{k=1 \\ U^{(a)}_{m_a,k}\neq 0}}^{K} \frac{\log(1+\frac{P^{(a)}_{m_a,k}H^{(a,a)}_{m_a}}{I^{(a)}_k+Z_k})}{P^{(a)}_{m_a,k}}, \quad (4.12)$$

where each $v^{(a)}_{m_a}$ can be any positive weight. At the BS a, let $S^{(a)}_{m_a,k} \triangleq \frac{P^{(a)}_{m_a,k}H^{(a,a)}_{m_a}}{I^{(a)}_k+Z_k}$ denote the SINR of session m_a on subchannel k.

Lemma 2 The unique solution of TPC strategy set that satisfies (4.7), i.e., $\tilde{\mathcal{P}}^N = \tilde{\mathcal{P}}^{(a)} \dot{\times} \tilde{\mathcal{P}}^{-a}$, maximizes $E^{(a)}$, subject to the constraints of Problem 2.

Proof We start from proving that $\tilde{\mathcal{P}}^N$ maximizes

$$\hat{E}^{(a)} \triangleq \sum_{m_a=1}^{M^{(a)}} v_{m_a}^{(a)} \sum_{\substack{k=1 \\ U_{m_a,k}^{(a)} \neq 0}}^{K} \frac{\log(1+S_{m_a,k}^{(a)})}{S_{m_a,k}^{(a)}}. \tag{4.13}$$

Define $f(s) = \frac{s}{1+s} - \log(1+s)$. It is trivial to prove that $f(s)$ is a non-increasing function of s. Hence, $f(S_{m_a,k}^{(a)}) \leq f(0) \equiv 0$, since each $S_{m_a,k}^{(a)} \geq 0$ always holds. Then, we have

$$\frac{\partial \hat{E}^{(a)}}{\partial S_{m_a,k}^{(a)}} = v_{m_a}^{(a)} \frac{\frac{S_{m_a,k}^{(a)}}{1+S_{m_a,k}^{(a)}} - \log(1+S_{m_a,k}^{(a)})}{(S_{m_a,k}^{(a)})^2} \leq 0 \text{ for } m_a \in \mathcal{M}^{(a)}; k \in \mathcal{K}. \tag{4.14}$$

As in the proof for Lemma 1, each $S_{m_a,k}^{(a)}$ is minimized to $U_{m_a,k}^{(a)} \gamma_{m_a}^{(a)}$ by $\tilde{\mathcal{P}}^N$. And $\hat{E}^{(a)}$ is maximized accordingly. To find the relationship between $E^{(a)}$ and $\hat{E}^{(a)}$, we can see that

$$\frac{\log(1+S_{m_a,k}^{(a)})}{P_{m_a,k}^{(a)}} = \frac{\log(1+S_{m_a,k}^{(a)})}{S_{m_a,k}^{(a)}} \frac{H_{m_a}^{(a,a)}}{I_k^{(a)} + Z_k} \text{ for } m_a \in \mathcal{M}^{(a)}; k \in \mathcal{K}. \tag{4.15}$$

On the RHS of (4.15), when each $\frac{\log(1+S_{m_a,k}^{(a)})}{S_{m_a,k}^{(a)}}$ is maximized by $\tilde{\mathcal{P}}^N$, the corresponding $I_k^{(a)}$ is minimized simultaneously. Hence, each $\frac{\log(1+S_{m_a,k}^{(a)})}{P_{m_a,k}^{(a)}}$ is maximized by $\tilde{\mathcal{P}}^N$ that solves (4.7), and so is the weighted sum $E^{(a)}$ with all positive weights. □

4.4.2 Multi-Cell TPC as Non-Cooperative Game

To study how each cell behaves in presence of other coexisting cells, we frame multi-cell TPC as a non-cooperative game, which serves as the second level of the URA game as in our two-level framework. In the TPC game, each cell solves Problem 2 independently.

Theorem 1 The unique solution of TPC strategy set that satisfies (4.7), i.e., $\tilde{\mathcal{P}}^N$, establishes the Nash equilibrium in the non-cooperative TPC game defined by Problem 2.

Proof According to the fixed point theorem [69] in game theory, two conditions must be satisfied for the existence of Nash equilibrium in the game: a) the strategy space, $\mathbb{P}^N = \dot{\times}_{n\in\mathcal{N}}\mathbb{P}^{(n)}$, for searching $\mathcal{P}^N$ should be a non-empty, compact, and convex subset of certain Euclidean space; b) the utility functions, $L^{(n)}$ for $n \in \mathcal{N}$, should be continuous in $\mathcal{P}^N$ and quasi-convex in $\mathcal{P}^{(n)}$. Due to the bounds defined in (4.3) for each $P^{(n)}_{m_n,k}$ in $\mathcal{P}^{(n)}$, $\mathbb{P}^{(n)}$ in each cell n is closed and bounded, and thus compact. It is also trivial to show that $\mathbb{P}^{(n)}$ is convex. We know that the Cartesian product of compact and convex sets is still compact and convex. Hence, the first condition is satisfied. As defined in (4.1), each $L^{(n)}$ is a linear combination of $P^{(n)}_{m_n,k}$ in $\mathcal{P}^{(n)}$ with all positive coefficients. Clearly, $L^{(n)}$ is continuous in $\mathcal{P}^N$. We know that a linear combination of convex functions with positive coefficients is again convex, and every convex function is also quasi-convex. Then, $L^{(n)}$ is quasi-convex in $\mathcal{P}^{(n)}$. Hence, the second condition is satisfied too.

As discussed in the proof for Lemma 1, not only $P^{(a)}_{m_a,k}$ in $\mathcal{P}^{(a)}$ but also $P^{(n)}_{m_n,k}$ in $\mathcal{P}^{-a}$ are minimized to $U^{(n)}_{m_n,k}\tilde{Q}^{(n)}_{m_n,k}$ by $\tilde{\mathcal{P}}^N$. These lower bounds are determined as soon as $\tilde{\mathcal{U}}^N$ is fixed. Regardless the different positive weights in coexisting cells' local objectives, the optimal TPC strategies applied in different cells are always the same unique one, i.e., $\tilde{\mathcal{P}}^N = \tilde{\mathcal{P}}^{(1)}\dot{\times}\tilde{\mathcal{P}}^{-1} = \cdots = \tilde{\mathcal{P}}^{(N)}\dot{\times}\tilde{\mathcal{P}}^{-N}$, that solves (4.7). According to the fixed point theorem, any fixed point of the best strategy space, $\mathbb{R}^N = \dot{\times}_{n\in\mathcal{N}}\mathbb{R}^{(n)}$, is the Nash equilibrium. Surely, the unique common best strategy set $\tilde{\mathcal{P}}^N$ is a fixed point of $\mathbb{R}^N$ in the TPC game, since any cell's best strategy is always to stay once $\tilde{\mathcal{P}}^N$ has been reached. As a result, we can prove that the Nash equilibrium is established by $\tilde{\mathcal{P}}^N$ in the TPC game. □

The Theorem 1 offers a guideline of designing TPC algorithms in a multi-cell system. In fact, there are already various existing TPC algorithms that do not rely on perfect inter-cell coordination, e.g., iterative water-filling algorithm [60].

4.5 Game of Subchannel Allocation

In this section, we focus on the SCA subproblem, and try to study multi-cell SCA as a non-cooperative game as well. However, unlike the case of TPC, several undesired facts for the SCA subproblem and game are briefly discussed before going any further.

4.5.1 Subchannel Allocation

If $\mathcal{P}^{(a)}$ and $\mathcal{P}^{-a}$ in Problem 1 are treated as known constants when selecting $\mathcal{U}^{(a)}$, the URA problem can be reduced to the SCA subproblem. For each selected $\mathcal{U}^{(a)}$ and fixed $\mathcal{U}^{-a}$, the setting of $\mathcal{P}^{(a)}$ and $\mathcal{P}^{-a}$ can be $\mathcal{P}^N = \mathcal{P}^{(a)} \dot{\times} \mathcal{P}^{-a} \equiv \tilde{\mathcal{P}}^N$, taking advantage of the unique optimal solution to Problem 2 as in Lemmas 1 and 2. Note that $\mathcal{P}^{(a)}$ and $\mathcal{P}^{-a}$ do not keep the same values every time when $\mathcal{U}^{(a)}$ is changed. Besides (4.7) for such setting, the conditions (4.5) for $a \in \mathcal{N}$, (4.6), and (4.8) for the setting of $\mathcal{U}^{(a)} \equiv \tilde{\mathcal{U}}^{(a)}$ in Problem 2 also become the constraints of the SCA subproblem in the cell a, which is defined as follows.

Problem 3 (SCA)

Given: $\mathcal{U}^{-a}$;

Find: $\mathcal{U}^{(a)}$;

Minimize: $L^{(a)}$;

Subject to: (4.5), (4.6), (4.7), (4.8).

In fact, solving Problem 1 is equivalent to solving Problem 3 once while solving Problem 2 multiple times (once for each change of SCA strategy when solving Problem 3). In our two-level framework, given each SCA strategy, say $\mathcal{U}^{(a)} \equiv \tilde{\mathcal{U}}^{(a)}$, taken by the cell a to solve Problem 3, Problem 2 is solved once to give $\tilde{\mathcal{P}}^N$ under $\mathcal{U}^{(a)} \equiv \tilde{\mathcal{U}}^{(a)}$. Then, the objective $L^{(a)}$ for this SCA strategy $\mathcal{U}^{(a)} \equiv \tilde{\mathcal{U}}^{(a)}$ can be computed under $\mathcal{P}^N \equiv \tilde{\mathcal{P}}^N$. Hence, $L^{(a)}$ can be regarded as a function of only $\mathcal{U}^{(a)}$ when always taking the optimal TPC strategy given each $\mathcal{U}^{(a)}$. Then, Problem 1 is reduced to Problem 3, which is a BIP.

4.5.2 Multi-Cell SCA as Non-Cooperative Game

Similar to the TPC game, we try to frame multi-cell SCA as a non-cooperative game, in which each cell solves Problem 3 independently. However, the SCA game, which serves as the first level of the URA game, does not have the good properties as the TPC game does.

Corollary 1 Solving Problem 3 for the optimal SCA strategy necessarily requires the perfect global knowledge of $\tilde{\mathcal{U}}^{-a}$, $\tilde{\mathcal{P}}^{-a}$, and $H_{m_n}^{(n,n')}$ for $n \in \mathcal{N}, n \neq a$ in the cell a.

Proof Obviously, real-time inter-cell coordination is necessary to keep track of $\tilde{\mathcal{U}}^{-a}$ and $\tilde{\mathcal{P}}^{-a}$. Although it is possible to estimate the values of $H_{m_a}^{(a,n')}$ in the cell a by decoding downlink pilot signals from BSs n' [56, 57], that of $H_{m_n}^{(n,n')}$ for $n \neq a$ are hard to get. □

Corollary 2 The complexity of searching for the optimal solution to Problem 3 is $O(K!)$.

Proof Due to the setting of $\mathcal{P}^N \equiv \tilde{\mathcal{P}}^N$, any session m_a's uplink capacity on each taken subchannel is always $\log(1+\gamma_{m_a}^{(a)})$ given $\gamma_{m_a}^{(a)}$. To minimize $L^{(a)}$ and make (4.8) hold, the number of subchannels taken by each session m_a is $T_{m_a}^{(a)} \triangleq \lceil \frac{\theta_{m_a}^{(a)}}{\log(1+\gamma_{m_a}^{(a)})} \rceil$. Then, the number of subchannels taken by the cell a is $T^{(a)} = \sum_{m_a=1}^{M^{(a)}} T_{m_a}^{(a)}$. In view of (4.5), the number of possible solutions for $\mathcal{U}^{(a)}$ is $\frac{K!}{(K-T^{(a)})!}$. Hence, the complexity of search via brute force is on the order of $O(K!)$. □

Corollary 3 The Nash equilibrium does not always exist in the non-cooperative SCA game defined by Problem 3.

Proof The discrete strategy space for searching $\mathcal{U}^{(a)}$ cannot satisfy the conditions for the existence of Nash equilibrium as in the proof for Theorem 1. □

4.6 Algorithms and Implementation

Towards the globally power-efficient and locally fair USFR, we propose a two-level game-theoretic approach. Instead of getting the optimal solution that relies on perfect global know-

ledge and high computational complexity, we resort to a heuristic yet practical approach.

4.6.1 Local URA Algorithm

In the cell a, the objective of solving Problem 3 is to find $\tilde{\mathcal{U}}^{(a)}$ that minimizes the weighted sum of sessions' power consumption. According to the user weights defined in (4.2), the numerator $\sum_{n=1,n\neq a}^{N} H_{m_a}^{(a,n)}$ characterizes session m_a's potential of interference *to* other cells, while the denominator $H_{m_a}^{(a,a)}$ characterizes session m_a's tolerance for interference *from* other cells. The sessions with larger weights are either more likely to cause interference or are more vulnerable to interference. Thus, based on the observations, such weights representing the sessions' priorities can be used to heuristically perform local SCA. Intuitively, near edge users tend to claim dedicated subchannels and need to be scheduled first, while inner users are the most coexistence-friendly ones and can be scheduled at last.

Now, we focus on a certain session, say m_a, that has been scheduled to take subchannels in the cell a. Suppose that the perfect knowledge of $\tilde{\mathcal{U}}^{-a}$, $\tilde{\mathcal{P}}^{-a}$, and $H_{m_n}^{(n,a)}$ for $n \neq a$ is not available. The only information that can be locally obtained by the BS a is $I_k^{(a)}$, i.e., the aggregated interference measurement on each subchannel k. Hence, the local SCA decisions for the session m_a is merely based on the interference measurements $\hat{I}_k^{(a)} \triangleq I_k^{(a)}(\tilde{\mathcal{P}}^N|_{U_{m_a,k}^{(a)}=0})$ or $\bar{I}_k^{(a)} \triangleq I_k^{(a)}(\tilde{\mathcal{P}}^N|_{U_{m_a,k}^{(a)}=1})$, where $\hat{I}_k^{(a)}$ ($\bar{I}_k^{(a)}$) is measured when no session (the session m_a) takes subchannel k and $\tilde{\mathcal{P}}^N$ is generated in the TPC game under the fixed $\tilde{\mathcal{U}}^{-a}$ and $U_{m_a,k}^{(a)} = 0$ ($U_{m_a,k}^{(a)} = 1$). Intuitively, the session m_a is supposed to take $T_{m_a}^{(a)}$ subchannels with the lowest $\bar{I}_k^{(a)}$ to meet the required SINR and QoS. However, the values of $\bar{I}_k^{(a)}$ are costly to get if inter-cell coordination is not assumed, since the session m_a has to try each subchannel k by triggering the TPC game to create $\bar{I}_k^{(a)}$. Instead, the session m_a can take $T_{m_a}^{(a)}$ subchannels with the lowest $\hat{I}_k^{(a)}$ thanks to the connection between $\hat{I}_k^{(a)}$ and $\bar{I}_k^{(a)}$. In most cases, $\bar{I}_{k_1}^{(a)} < \bar{I}_{k_2}^{(a)}$ is true if $\hat{I}_{k_1}^{(a)} < \hat{I}_{k_2}^{(a)}$ for the same session m_a. Therefore, a heuristic solution of SCA strategy to Problem 3 can be given by Algorithm 1.

Algorithm 1 Local URA algorithm

1: **sort** the sequence of sessions m_a for $m_a \in \mathcal{M}^{(a)}$ by $w_{m_a}^{(a)}$ in descending order, and store the sorted sequence in array $\mathbf{M}(i)$ for $i = 1 \to M^{(a)}$
2: **set** $\mathcal{U}^{(a)} = \{\mathbf{0}\}$, and measure $\hat{I}_k^{(a)}$ for $k \in \mathcal{K}$ generated in the TPC game
3: **sort** the sequence of subchannel k for $k \in \mathcal{K}$ by $\hat{I}_k^{(a)}$ in ascending order, and store the sorted sequence in array $\mathbf{K}(j)$ for $j = 1 \to K$
4: **for** $i = 1 \to M^{(a)}$ **do**
5: **set** $m_a = \mathbf{M}(i)$
6: **for** $j = 1 \to K$ **do**
7: **set** $k = \mathbf{K}(j)$
8: **if** $\sum_{m=1}^{M^{(a)}} U_{m,k}^{(a)} == 0$ **then**
9: **set** $U_{m_a,k}^{(a)} = 1, \cdots, U_{m_a,k+T_{m_a}^{(a)}-1}^{(a)} = 1$
10: **break**
11: **end if**
12: **end for**
13: **end for**
14: **measure** $\bar{I}_k^{(a)}$ for k whose $\sum_{m=1}^{M^{(a)}} U_{m,k}^{(a)} == 1$ generated in the TPC game, and adjust $P_{m_a,k}^{(a)}$ to meet the required SINR and QoS if $U_{m_a,k}^{(a)} == 1$

Corollary 4 The complexity of running Algorithm 1 is $O(KM^{(a)})$.

Proof The major cost for running Algorithm 1 comes from the two sorting operations in lines 1 and 3 and the nested FOR loops between lines 4 and 13. Sorting sessions leads to the complexity of $O(M^{(a)} \log M^{(a)})$ via e.g., merge sort. Likewise, sorting subchannels incurs $O(K \log K)$. The nested FOR loops result in $O(KM^{(a)})$. Because $K > M^{(a)}$ and possibly $M^{(a)} > \log K$, the complexity of running Algorithm 1 is on the order of $O(KM^{(a)})$. □

4.6.2 Two-Level Game-Theoretic Algorithm

Due to the fact that the SCA game does not always converge, a cost function [69] should be considered in the definition of utility. We start from a cooperative URA game where

coexisting cells share the same global objective, G, which can be defined as

$$G \triangleq \sum_{n=1}^{N} L^{(n)} = \sum_{n=1}^{N} \sum_{m_n=1}^{M^{(n)}} w_{m_n}^{(n)} \sum_{k=1}^{K} P_{m_n,k}^{(n)}, \tag{4.16}$$

which is the weighted sum of sessions' power consumption in all cells. It is trivial to show that the cooperative game in which each cell solves Problem 1 with common global objective G can converge to the globally optimal pattern with unique optimal value of G. We can rewrite (4.16) as $G = L^{(a)} + C_g^{(a)}$, where the cell a's global cost function can be rewritten as

$$C_g^{(a)} \triangleq \sum_{n=1,n\neq a}^{N} \sum_{m_n=1}^{M^{(n)}} \sum_{n'=1,n'\neq n}^{N} \sum_{k=1}^{K} \frac{P_{m_n,k}^{(n)} H_{m_n}^{(n,n')}}{H_{m_n}^{(n,n)}}. \tag{4.17}$$

On the RHS of (4.17), each $P_{m_n,k}^{(n)} H_{m_n}^{(n,n')}$ represents the interference from session m_n to BS n'. In the cell a, however, only the sum of $P_{m_n,k}^{(n)} H_{m_n}^{(n,a)}$ on each subchannel k, i.e., $I_k^{(a)}$, is known. Hence, the cell a's local cost function, denoted by $C_l^{(a)}$, can be defined as

$$C_l^{(a)} \triangleq \delta^{(a)} \sum_{k=1}^{K} I_k^{(a)}, \tag{4.18}$$

where $\delta^{(a)}$ denotes a positive price factor. Note that unlike the commonly defined cost functions, such as $\sum_{m_a=1}^{M^{(a)}} \sum_{k=1}^{K} \log(P_{m_a,k}^{(a)})$ [69], that change in the opposite direction of the change in $L^{(a)}$ while adapting any $P_{m_a,k}^{(a)}$, the defined $C_l^{(a)}$ and $L^{(a)}$ can be reduced together most of the time. When the cell a performs local URA to minimize $L^{(a)}$, most of the coexisting cells in the multi-cell system can reduce power simultaneously for the drop of interference from the cell a.

After replacing $L^{(a)}$ with $L^{(a)} + C_l^{(a)}$ in Problems 2 and 3, we get the revised TPC and SCA subproblems in the cell a as Problems 4 and 5, respectively.

Lemma 3 Given that the previously fixed SCA strategy set $\tilde{\mathcal{U}}^N = \tilde{\mathcal{U}}^{(a)} \dot{\times} \tilde{\mathcal{U}}^{-a}$ satisfies (4.5) for $a \in \mathcal{N}$, (4.6), and (4.8), the unique optimal solution of TPC strategy set, say

$\tilde{\mathcal{P}}^N = \tilde{\mathcal{P}}^{(a)} \dot{\times} \tilde{\mathcal{P}}^{-a}$, to Problem 4 satisfies (4.7). Globally, $\tilde{\mathcal{P}}^N$ further establishes the Nash equilibrium in the non-cooperative TPC game defined by Problem 4.

Proof The proof follows the same logic as that for Lemma 1 and Theorem 1. □

As above, solving the revised Problem 1 with objective $L^{(a)} + C_l^{(a)}$ is equivalent to solving Problem 5 once while solving Problem 4 multiple times (once for each change of SCA strategy when solving Problem 5). To guarantee the convergence of the URA game to a commonly agreed coexistence pattern, the adaptation of $\delta^{(n)}$ in each $C_l^{(n)}$ is necessary to eliminate the interest conflicts among greedy cells in the game. We do not assume available perfect global knowledge and centralized decision maker like virtual referee [62]. Intuitively, it is difficult to ensure that the non-cooperative URA game defined by $C_l^{(a)}$ always converges to the so-called globally optimal coexistence pattern created by the cooperative URA game defined by $C_g^{(a)}$. Moreover, even such an optimal coexistence pattern can be established via perfect inter-cell coordination, it can be subject to frequent change due to dynamic spectrum environment and potential user mobility. Combined with Algorithm 1, a heuristic yet distributed Algorithm 2 is proposed. We next prove that the cells running Algorithm 2 independently can converge to a common stable coexistence pattern without inter-cell coordination.

In Algorithm 2, the loop between lines 2 and 17 describes the cell a's behavior in the SCA game. Whenever the cell a takes a new SCA strategy, the TPC game as in lines 2 and 14 in Algorithm 1 is played and achieves the Nash equilibrium as in Lemma 3. The use of backoff timer between lines 3 and 6 aims to simplify the scheduling of acting cells in the multi-cell system. The time interval for the backoff timer can be set in seconds. When the time for solving the URA problem is relatively short, the coexisting cells are expected to take turns to act in the URA game. Then, the non-cooperative URA game is achieved by successive local searches of SCA and TPC strategies optimizing local objectives. The adaptation of $\delta^{(a)}$ in line 15 is to make the cell a contribute to the Nash equilibrium in the URA game.

Algorithm 2 Two-level game-theoretic algorithm

set $\delta^{(a)} = \delta_0^{(a)}$ ($\delta_0^{(a)}$ is a positive constant), and begin to participate in the SCA game
loop
  set the backoff timer for a random time interval
  repeat
    do the countdown of backoff timer
  until the backoff timer expires
  measure $\tilde{I}_k^{(a)}$ for $k \in \mathcal{K}$, and compute $\tilde{L}^{(a)}$ under the current $\mathcal{U}^N$ and $\tilde{\mathcal{P}}^N$ generated in the TPC game
  set $L_0^{(a)} = \tilde{L}^{(a)}$, $I_0^{(a)} = \sum_{k=1}^K \tilde{I}_k^{(a)}$, and $F_0^{(a)} = L_0^{(a)} + \delta_0^{(a)} I_0^{(a)}$, and record $\mathcal{U}_0^{(a)} = \mathcal{U}^{(a)}$
  run Algorithm 1
  set $L^{(a)} = \tilde{L}^{(a)}$, $I^{(a)} = \sum_{k=1}^K \tilde{I}_k^{(a)}$, and $F^{(a)} = L^{(a)} + \delta^{(a)} I^{(a)}$ under the updated $\mathcal{U}^N$ and $\tilde{\mathcal{P}}^N$ from Algorithm 1
  if $F^{(a)} \geq F_0^{(a)}$ **then**
    set $\mathcal{U}^{(a)} = \mathcal{U}_0^{(a)}$ (give up acting in the SCA game)
  end if
  if $!(L^{(a)} \leq L_0^{(a)}$ && $I^{(a)} \leq I_0^{(a)})$ **then**
    update $\delta^{(a)} = \delta^{(a)} + \Delta^{(a)}$ ($\Delta^{(a)}$ is a positive constant)
  end if
end loop

Theorem 2 In the non-cooperative URA game defined by Problems 4 and 5, the cells that run Algorithm 2 independently agree on a common stable coexistence pattern under the established Nash equilibrium without inter-cell coordination.

Proof As in Lemma 3, the TPC game defined by Problem 4 can always achieve the Nash equilibrium whenever any cell in the game changes its SCA strategy. Then, we verify that the cells running Algorithm 2 can agree on a common $\tilde{\mathcal{U}}_0^N = \dot{\times}_{n \in \mathcal{N}} \tilde{\mathcal{U}}_0^{(n)}$ in the SCA game, and nobody has an intention of making changes to the converged $\tilde{\mathcal{U}}_0^N$.

According to the price adaptation strategy in Algorithm 2, $\delta^{(a)}$ in the cell a is increased by $\Delta^{(a)}$ if the taken SCA strategy does not contribute to the desired Nash equilibrium. The cell a contributes to the Nash equilibrium only when $L^{(a)} \leq L_0^{(a)}$ and $I^{(a)} \leq I_0^{(a)}$. Hence, $\delta^{(a)}$ is non-decreasing. The cell a changes subchannel allocation only when $F^{(a)} < F_0^{(a)}$. Hence, $F_0^{(a)}$ is decreasing. If $\delta^{(a)}$ has been increased to always make $F^{(a)} \geq F_0^{(a)}$ hold, the cell a no longer takes any action in the SCA game. Then, the value of $F_0^{(a)}$ keeps unchanged, say

$F_0^{(a)} \equiv \tilde{F}_0^{(a)}$. The adaptation of $\delta^{(n)}$ for $n \neq a$ is similar. Once all the cells give up acting in the SCA game, a certain common $\tilde{\mathcal{U}}_0^N$ is agreed on. In the end, the Nash equilibrium is established by the TPC game under the fixed $\tilde{\mathcal{U}}_0^N$.

Now we prove that the achieved common coexistence pattern is stable. The Lyapunov's direct stability theorem in nonlinear control theory: if there exists a Lyapunov function on a region containing a critical point, which is set as the origin, then the origin is a stable critical point and all solutions originating in the region approach the origin asymptotically [71]. At the convergence point, $F_0^{(n)} \equiv \tilde{F}_0^{(n)}$ for $n \in \mathcal{N}$. Hence, the converged coexistence pattern is a critical point. We can further show that $\sum_{n=1}^{N}(F_0^{(n)} - \tilde{F}_0^{(n)})$ is a Lyapunov function. If $\sum_{n=1}^{N}(F_0^{(n)} - \tilde{F}_0^{(n)})$ is a Lyapunov function, it needs to satisfy the following three conditions:

1) The summation $\sum_{n=1}^{N}(F_0^{(n)} - \tilde{F}_0^{(n)}) = 0$ at the origin, i.e., the convergence point, and $\sum_{n=1}^{N}(F_0^{(n)} - \tilde{F}_0^{(n)}) > 0$ for other coexistence patterns;

2) The derivative $\frac{d\sum_{n=1}^{N} F_0^{(n)}}{d\mathbf{p}} > 0$, where $\mathbf{p}$ is any vector radiating from the origin;

3) The derivative $\frac{d\sum_{n=1}^{N} F_0^{(n)}}{dt} < 0$, where t is time.

We know that $\sum_{n=1}^{N} F_0^{(n)}$ reaches its lower bound $\sum_{n=1}^{N} \tilde{F}_0^{(n)}$ at the convergence point. Hence, condition 1) is satisfied. The adaptation of $\delta^{(n)}$ for $n \in \mathcal{N}$ can make $\sum_{n=1}^{N} F_0^{(n)}$ decreasing, so the convergence point is the lowest point in terms of $\sum_{n=1}^{N} F_0^{(n)}$. Hence, condition 2) is satisfied. We know that $\sum_{n=1}^{N} F_0^{(n)}$ is decreasing in time. Hence, condition 3) is satisfied. Then, $\sum_{n=1}^{N}(F_0^{(n)} - \tilde{F}_0^{(n)})$ is a Lyapunov function. Based on the Lyapunov's direct stability theorem, we conclude that the converged common coexistence pattern is stable. □

The converged coexistence pattern is achieved by successive local searches of SCA and TPC strategies optimizing local objectives. We have shown that the solution to the local SCA subproblem may not be optimal without having perfect global knowledge and requiring high computational complexity. Furthermore, successive local searches do not guarantee the resulting coexistence pattern to be unique and globally optimal. Hence, the proposed

approach as in Algorithm 2 is a heuristic. However, our approach does not rely on any inter-cell coordination and has low complexity. It is always based on the actual local interference measurements and does not need any estimates of environment. In section 4.7, we show that the converged coexistence pattern in the non-cooperative URA game is usually close enough to the globally optimal coexistence pattern obtained in the cooperative URA game.

4.6.3 Implementation Variants

So far, our discussion of USFR was based on the definition of user weights as in (4.2) and the assumption that coexisting networks belong to different network operators. In fact, however, there are some other implementation variants of USFR that can be considered.

Variant 1 The first variant of USFR technique is obtained by modifying the definition of $w_{m_a}^{(a)}$ in Algorithm 1, which is not limited to (4.2) and can be set on demand. The purpose of optimizing the weighted sum of sessions' power consumption or power efficiency is to enhance intra-cell fairness locally as well as spectrum utilization globally.

Here, we present two alternate definitions for the user weight $w_{m_a}^{(a)}$. The first definition reflects the fact that user terminals in each cell may not be equal in terms of spectrum access priority. For instance, some users may pay a higher price for higher priority access. If such prioritized spectrum access is considered, $w_{m_a}^{(a)}$ can be defined as

$$w_{m_a}^{(a)} \triangleq \alpha p_{m_a}^{(a)} + (1-\alpha)\bar{w}_{m_a}^{(a)}, \tag{4.19}$$

where $p_{m_a}^{(a)}$ denotes session m_a's predefined normalized priority; $\bar{w}_{m_a}^{(a)}$ denotes session m_a's normalized weight based on (4.2); and α is a convex combination factor.

The second definition reflects the fact that a BS, say a, may have difficulty in measuring $\sum_{n=1,n\neq a}^{N} H_{m_a}^{(a,n)}$ locally for each session m_a, since the values of $H_{m_a}^{(a,n)}$ for mobile/portable users and far-away BSs may be very difficult to calculate accurately. Because precise values of $w_{m_a}^{(a)}$ are not necessarily required for uplink scheduling, an approximation of $\sum_{n=1,n\neq a}^{N} H_{m_a}^{(a,n)}$

can be used instead. Then, $w_{m_a}^{(a)}$ in terms of worst-case interference can be defined as

$$w_{m_a}^{(a)} \triangleq \frac{\bar{H}_{m_a}^{(a,n)}}{H_{m_a}^{(a,a)}}, \tag{4.20}$$

where $\bar{H}_{m_a}^{(a,n)} = \max_{n \in \mathcal{N}, n \neq a} \{H_{m_a}^{(a,n)}\}$, or $\bar{H}_{m_a}^{(a,n)}$ is the propagation gain from session m_a to the nearest BS n if propagation gain is inversely proportional to distance. Moreover, if $\bar{H}_{m_a}^{(a,n)} = 1$, the definition is reduced to the way of distinguishing inner and edge users in the downlink.

Variant 2 The second variant of USFR technique is the application to a single-operator coexistence scenario. In our previous problem formulation, we assumed a multi-operator case where each coexisting cell is controlled by a distinct network operator. Here, we assume that a single network operator centrally manages all the N co-channel cells and $M = \sum_{n=1}^{N} M^{(n)}$ uplink sessions. In such a case, in addition to SCA and TPC, user-cell association (UCA) needs to be considered. Each session m for $m \in \mathcal{M} \triangleq \{1, \cdots, M\}$ needs to decide which cell to associate with. Hence, an extra set of binary variables $\mathcal{R}^{(n)} \triangleq \{R_m^{(n)}; m \in \mathcal{M}\}$ is used to characterize the association of sessions in each cell n. Each $R_m^{(n)}$ is equal to 1 (0) when session m associates (does not associate) with cell n.

In this scenario, multi-cell URA can be formulated as a centralized optimization problem with the global objective G defined in (4.16). Based on (4.3), (4.4), and (4.5), the constraints for this problem are as follows. First, each $P_{m,k}^{(n)}$ should be bounded. We have

$$R_m^{(n)} U_{m,k}^{(n)} \hat{Q}_{m,k}^{(n)} \leq P_{m,k}^{(n)} \leq R_m^{(n)} U_{m,k}^{(n)} \bar{Q}_{m,k}^{(n)} \text{ for } n \in \mathcal{N}; m \in \mathcal{M}; k \in \mathcal{K}. \tag{4.21}$$

Second, each session m's QoS requirement should be met, and hence we have

$$\sum_{n=1}^{N} \sum_{k=1}^{K} \log(1 + \frac{R_m^{(n)} P_{m,k}^{(n)} H_m^{(n,n)}}{I_k^{(n)} + Z_k}) \geq \theta_m^{(n)} \text{ for } m \in \mathcal{M}. \tag{4.22}$$

Third, each subchannel k should be allocated to no more than one session in cell n. We have

$$\sum_{m=1}^{M} R_m^{(n)} U_{m,k}^{(n)} \leq 1 \text{ for } n \in \mathcal{N}; k \in \mathcal{K}. \tag{4.23}$$

Fourth, each session m should only be associated with one cell, and hence we have

$$\sum_{n=1}^{N} R_m^{(n)} = 1 \text{ for } m \in \mathcal{M}. \tag{4.24}$$

Hence, the centralized single-operator optimization problem (SOP) is defined as follows.

Problem 6 (SOP)

Find: $\mathcal{R}^{(n)}, \mathcal{U}^{(n)}, \mathcal{P}^{(n)}$ for $n \in \mathcal{N}$;

Minimize: G;

Subject to: (4.21), (4.22), (4.23), (4.24).

Solving Problem 6 needs to determine $\mathcal{R}^{(n)}$. Fortunately, the values of $\mathcal{R}^{(n)}$ can be fixed in advance to simplify Problem 6 as the following corollary prescribes.

Corollary 5 Each session m always gets associated with the cell $n = \arg\max_{n \in \mathcal{N}} \{H_m^{(n,n)}\}$ for the minimization of G, in which each $w_m^{(n)}$ is defined in (4.2).

Proof To minimize G, each session m's $w_m^{(n)}$ needs to be minimized. When a session m registers to the cell $n = \arg\max_{n \in \mathcal{N}} \{H_m^{(n,n)}\}$, we have $\min_{n \in \mathcal{N}} w_m^{(n)} = \frac{\sum_{n=1}^{N} H_m^{(n,n)} - \max_{n \in \mathcal{N}} \{H_m^{(n,n)}\}}{\max_{n \in \mathcal{N}} \{H_m^{(n,n)}\}}$. □

Therefore, whenever a session becomes active, it first detects downlink signals from nearby BSs and then associates with the cell n with $\max_{n \in \mathcal{N}} \{H_m^{(n,n)}\}$ or simply the nearest cell. After the values of $\mathcal{R}^{(n)}$ are fixed, we can solve Problem 6 by the previously defined cooperative URA game, where our approach is still applicable. Note that USFR is still decentralized in the single-operator case. Although a centralized optimization can be conducted by the operator, our approach does not require any extra infrastructure and high complexity.

4.7 Performance Evaluation

In this section, real-time performance of USFR achieved by Algorithm 2 is evaluated. We focus on a self-coexistence scenario, as in Figure 4.2, where $N = 7$ CR network cells are forced to share a single TV white space channel. Each cell is centrally controlled by a BS, and direct inter-cell coordination is not needed. In each cell n, $M^{(n)}$ uplink sessions are generated and randomly distributed. Through simulations, we evaluate a number of different definitions for the user weight: D0 as defined in (4.2), D1 as defined in (4.19), D2 as defined in (4.20) with $\bar{H}_{m_a}^{(a,n)} = \max_{n\in\mathcal{N},n\neq a}\{H_{m_a}^{(a,n)}\}$, and D3 as defined in (4.20) with $\bar{H}_{m_a}^{(a,n)} = 1$. We also apply Corollary 5 to each multi-operator (MO) scenario, and study the corresponding single-operator (SO) case. The average cell radius is 4 km, and the propagation gains are computed via log-distance path-loss model with exponent 3.2. The other parameters are set as follows: $\gamma_{m_n}^{(n)} = 30$, $T_{m_n}^{(n)} = 1$, $\bar{Q}_{m_n,k}^{(n)} = 100$ mW, Z_k is set by noise density -174 dBm/Hz and TV channel width 6 MHz. The coexistence patterns created by Algorithm 2 under different combinations of total number of sessions $M \triangleq \sum_{n=1}^{7} M^{(n)}$ and degree of spectrum sharing $\kappa \triangleq \frac{K}{M}$ are analyzed. In the following paragraphs, we discuss the various attributes of USFR based on our simulation experiment results.

First, we evaluate the convergence speed of USFR. The empirical cumulative distribution function (ECDF) of the number of iterations for converging to a common coexistence pattern is shown in Figure 4.3. Here, one iteration represents a change of URA strategy by any cell in the 7-cell system. The starting point of $\mathcal{U}_0^N$ in Algorithm 2 is any feasible point to Problem 1. We can see that the convergence speed of USFR is relatively fast even when the shared channel is heavily utilized.

Second, we evaluate the spectrum utilization efficiency of USFR. The distributed USFR is compared with the centralized frequency planning in Figure 4.4 in terms of the minimum K required to accommodate M sessions. With the global knowledge of the 7-cell topology as in Figure 4.2, the frequency planning typically requires $\frac{3M}{7}$ subchannels to accommodate M sessions with $T_{m_n}^{(n)} = 1$. This is because the cells 2, 4, and 6 can share the same subchannels

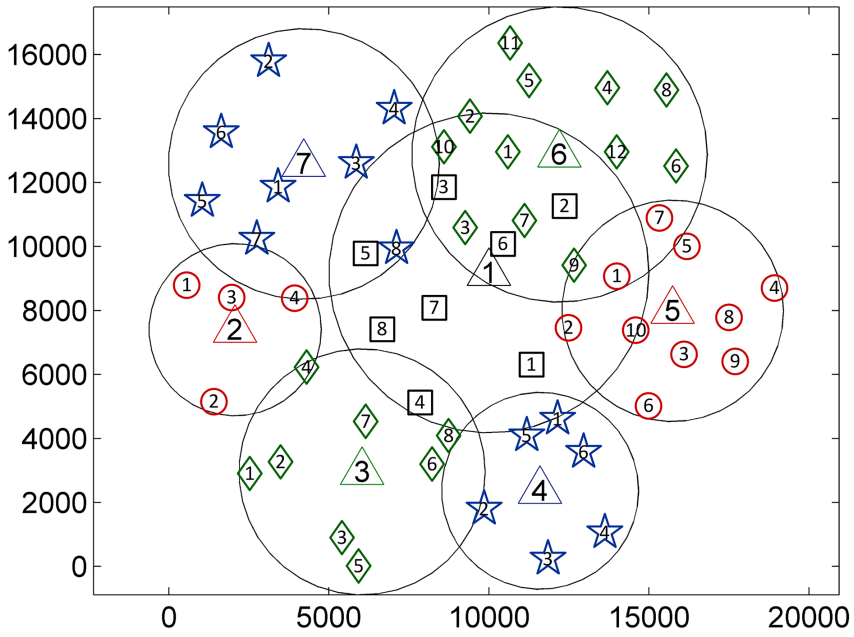

Figure 4.2: A 7-cell simulation scenario (triangles: BSs; other shapes: sessions; numbers: IDs of BSs and sessions; X/Y coordinates are in meters).

safely, and so can the cells 3, 5, and 7. We can see that our USFR under various definitions of user weights generally achieves a similar level of spectrum utilization without inter-cell coordination. In each cell, most inner users can coexist with edge users, probably far edge users, in other cells with acceptable mutual interference. User weight definition D0 outperforms the other definitions due to the use of the greater amount of information on the spectrum environment. Nonetheless, D2 and D3 achieve reasonable spectrum utilization. For the same number of subchannels, from Figure 4.4, we can also get the maximum number of sessions that can always be supported by the seven cells. If the number of sessions goes beyond the maximum limit, a feasible solution to the URA problem may not be guaranteed. Hence, some sessions have to be blocked if M is too large for a certain K. The blocking strategy depends on the cells' locations and coverage ranges, and the sessions' density, distribution, priority and fairness, etc. Note that the user weights defined in (4.2) should not be used to block users or sessions, since the weights only determine the quality of granted subchannels in a feasible solution to the URA problem. A session with a low weight value still has access

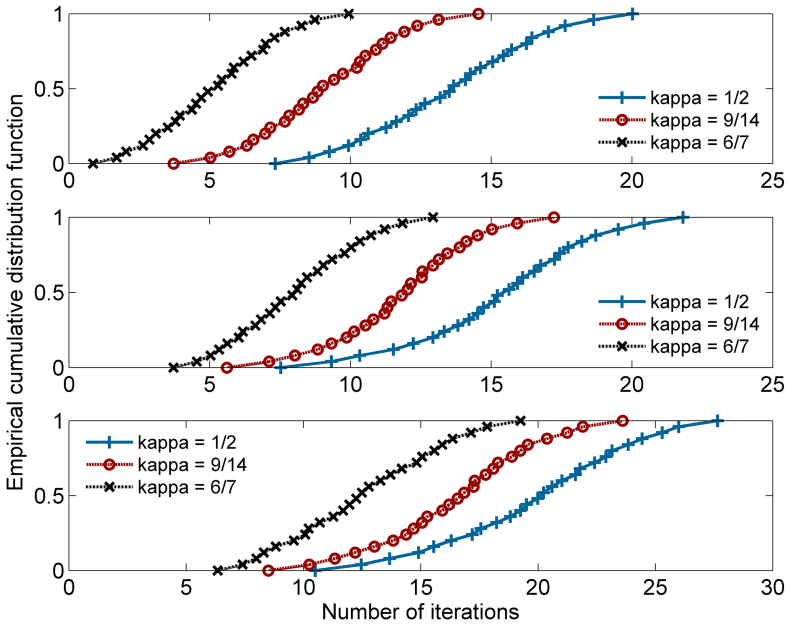

Figure 4.3: Convergence speed of USFR technique ($\Delta = 1 \times 10^{10}$): a) $M = 28$ (top); b) $M = 56$ (middle); c) $M = 112$ (bottom).

to low-quality subchannels in a feasible solution. Another type of priority should be used to make an unfeasible solution feasible by blocking users or sessions.

Third, we evaluate the power consumption and power efficiency of USFR. The objective ratio $\frac{\textit{heuristic}\text{ value of }G}{\textit{optimal}\text{ value of }G}$ is plotted in Figure 4.5. The heuristic G is computed by using the common coexistence pattern achieved in the non-cooperative URA game under $C_l^{(n)}$, while the optimal G is computed by using the globally optimal coexistence pattern obtained in the cooperative URA game under $C_g^{(n)}$. On the one hand, the heuristic coexistence pattern is generated by successive local searches of SCA and TPC strategies optimizing local objectives, so different orders of cell scheduling may result in different heuristic values of G. In Algorithm 2, we employ random cell scheduling to avoid inter-cell coordination. On the other hand, the optimal coexistence pattern is created by coordinating the cells to optimize the same global objective. The optimal value of G is unique, which cannot be improved by any cell via inter-cell coordination. Both heuristic and optimal coexistence patterns have the same format, which includes $\mathcal{U}^{(n)}$ and $\mathcal{P}^{(n)}$ for $n \in \mathcal{N}$, and G is a function of $\mathcal{U}^{(n)}$ and

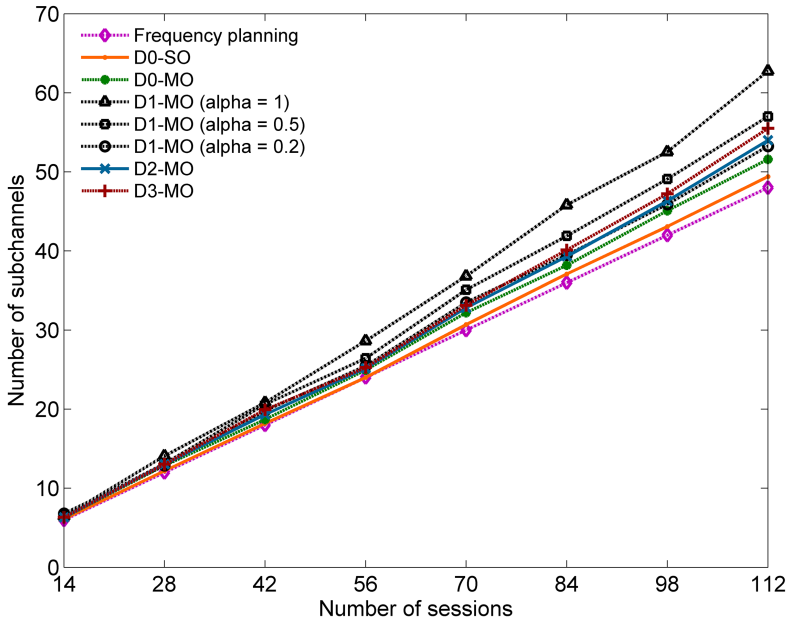

Figure 4.4: Spectrum utilization of USFR technique.

$\mathcal{P}^{(n)}$, so we can compare the heuristic solutions with the optimal solution. From Figure 4.5, we can see that our USFR achieves near-optimal performance in most cases, especially when the shared channel is not over saturated. The ratio $\frac{\text{average heuristic value of } G \text{ for SO}}{\text{average heuristic value of } G \text{ for MO}}$ is shown in Figure 4.6. We can see a significant advantage in minimizing G if the seven cells are controlled by one operator. This is due to the reduction of the cells' RF footprints according to Corollary 5. In Figure 4.7, the ratios $\frac{\text{average power efficiency for D1/D2/D3}}{\text{average power efficiency for D0}}$ are compared. For a fair comparison, the power efficiency (without user weights) is defined by the ratio of the sum of sessions' uplink capacity to the sum of sessions' transmit power. We can see that D1 with a small value of α, D2, and D3 achieve similar power efficiency.

Finally, we evaluate the intra-cell fairness of USFR. Firstly, we study a specific scenario as illustrated in Figure 4.2. The sessions' received power levels at their home BSs are shown in Figure 4.8, in which each set of bars represents the received power levels at a certain BS and the bars in each set are ordered by sessions' ID numbers and distinguished by different colors. As we can see, the received power levels at the BS are generally the same regardless

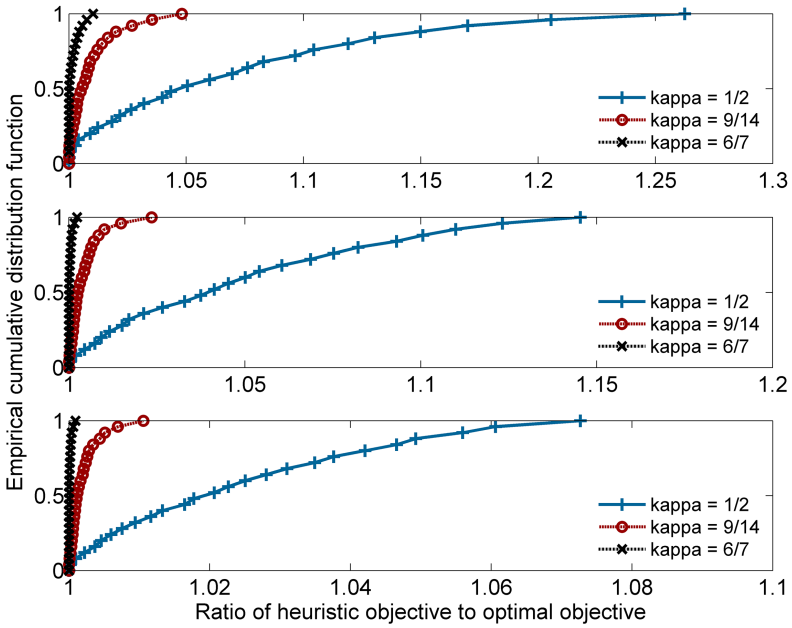

Figure 4.5: Power consumption of USFR technique (multi-operator): a) $M = 28$ (top); b) $M = 56$ (middle); c) $M = 112$ (bottom).

of the sessions' link distances. The near edge users (that are associated with sessions 2, 3, 4, and 5) in cell 1 tend to utilize subchannels exclusively. However, the inner users (that are associated with sessions 6 and 7) in cell 1 have to coexist with users in other cells, and thus need to receive signals at a higher received power to guarantee the minimum required SINR. Next, we focus on a general case. In each cell n, the fairness factor is defined by $\frac{\max_{m_n \in \mathcal{M}^{(n)}, k \in \mathcal{K}, U^{(n)}_{m_n,k} \neq 0} \{P^{(n)}_{m_n,k} H^{(n,n)}_{m_n}\}}{\min_{m_n \in \mathcal{M}^{(n)}, k \in \mathcal{K}, U^{(n)}_{m_n,k} \neq 0} \{P^{(n)}_{m_n,k} H^{(n,n)}_{m_n}\}}$, as shown in Figure 4.9. We can see that the proposed USFR technique always tries to reduce the fairness factor as more subchannels become available, and thus improves intra-cell fairness.

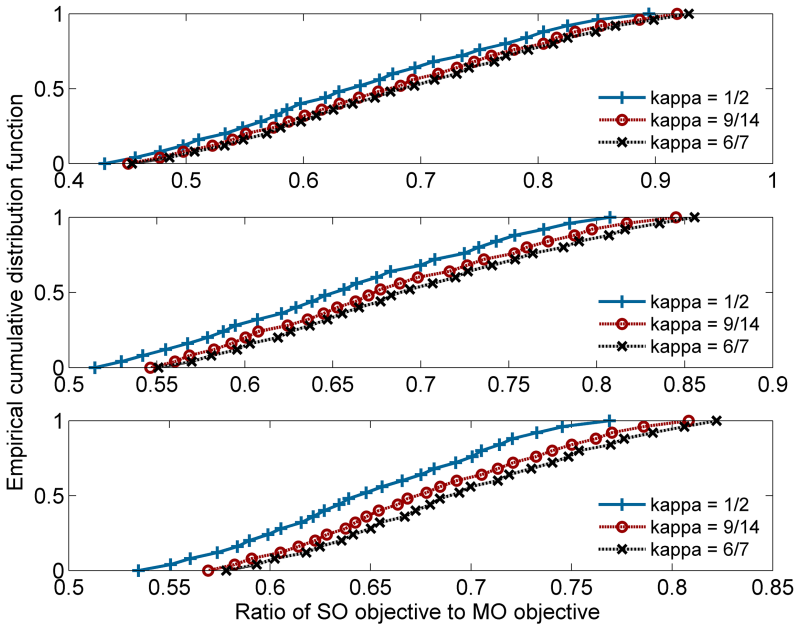

Figure 4.6: Power consumption of USFR technique (single-operator): a) $M = 28$ (top); b) $M = 56$ (middle); c) $M = 112$ (bottom).

4.8 Chapter Summary

In this chapter, the technique of USFR has been proposed to enhance the self-coexistence of CR networks. In view of the distributed nature of CR networks in a dynamic spectrum environment, we have resorted to the two-level game-theoretic approach that does not rely on direct inter-network coordination. In the non-cooperative game for multi-cell URA, the second-level TPC is solved optimally, while the first-level SCA is done heuristically in avoidance of perfect global knowledge and high computational complexity. Based on a well designed heuristic, the proposed USFR technique has been shown to effectively improve self-coexistence in spectrum utilization, power consumption, and intra-cell fairness.

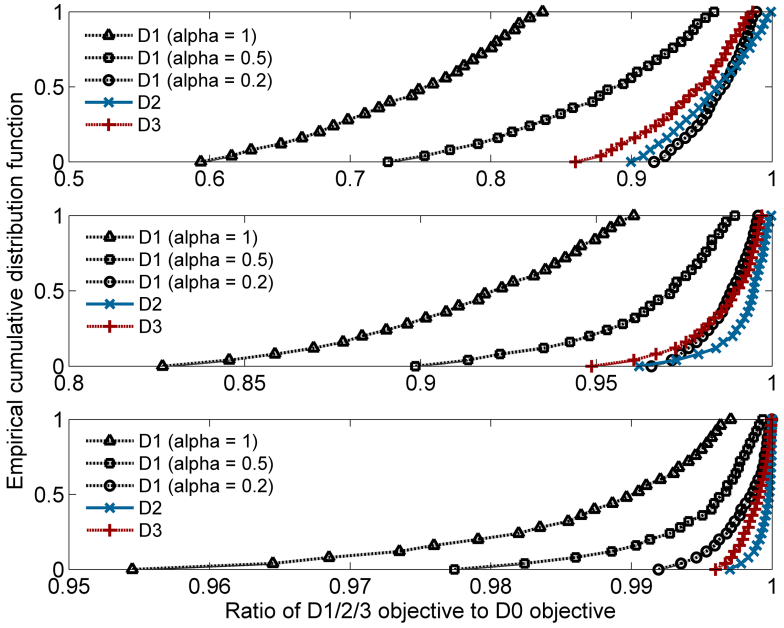

Figure 4.7: Power efficiency of USFR technique ($M = 56$): a) $\kappa = 1/2$ (top); b) $\kappa = 9/14$ (middle); c) $\kappa = 6/7$ (bottom).

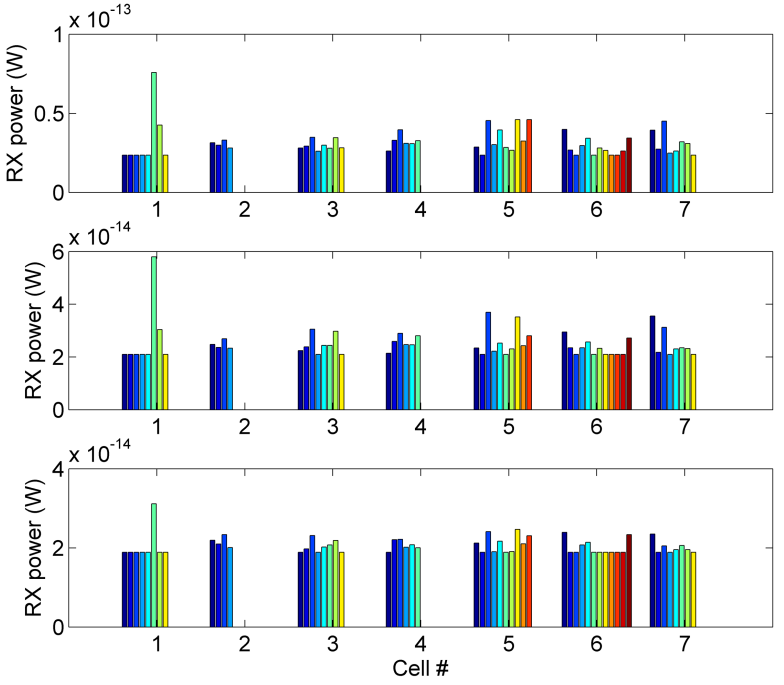

Figure 4.8: A sample of received power levels at home BS ($M = 56$): a) $\kappa = 4/7$ (top); b) $\kappa = 9/14$ (middle); c) $\kappa = 5/7$ (bottom).

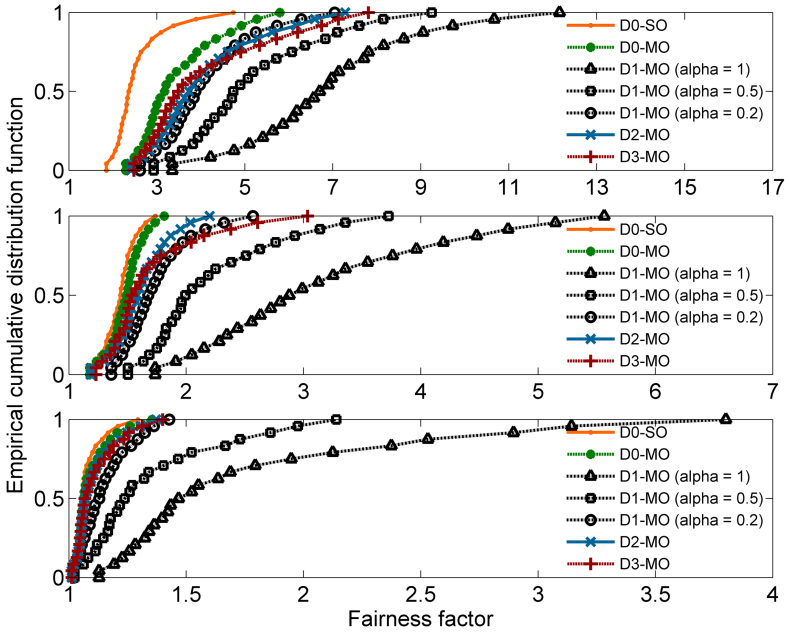

Figure 4.9: Intra-cell fairness of USFR technique ($M = 56$): a) $\kappa = 1/2$ (top); b) $\kappa = 9/14$ (middle); c) $\kappa = 6/7$ (bottom).

Chapter 5

Credit-Token-Based Spectrum Etiquette for Coexistence of Heterogeneous CR Networks

In this chapter, we focus on the third scenario to study the coexistence of heterogeneous CR networks, when co-channel networks use different air interface standards [17].

5.1 Challenges and Contributions

The increasing level of congestion in wireless spectrum has spurred a growing interest in CR technology. One major challenge that needs to be addressed in OSA is the potential interference from secondary users to primary users. For the most part, the problem of incumbent protection has been tackled by the use of incumbent geolocation databases augmented with spectrum sensing techniques. In contrast, however, the coexistence of secondary CR networks with equal priority, especially in the context of heterogeneous coexistence [8], is still a new area. We predict that the problem of network coexistence will garner great attention

in the near future, as the proliferation of white space networks becomes closer to reality, especially in dense urban environments. Recently, IEEE 802.19.1 [15] and a COGEU project [16] are being developed to provide general infrastructure-based solutions to the coexistence of dissimilar networks in TVWSs. However, employing extra coexistence infrastructures involves significant deployment and operational costs.

We are interested in a coexistence scenario where a centralized spectrum manager is not available. We propose a *credit-token*-based spectrum etiquette (CTSE) framework for the coexistence of heterogeneous CR networks with equal priority. The framework was inspired by the credit-token-based coexistence protocol (CT-CXP) in IEEE 802.16h [11]. Specifically, a two-stage *game-auction* coexistence framework is proposed. At the first stage, coexisting requesting networks (requesters) compete with each other as players in a non-cooperative game to obtain extra access to under-utilized white-space spectrum from co-located offering networks (offerers). At the second stage, each offerer rents out surplus spectrum to requesters through a truthful auction and allows spectrum sharing among multiple auction winners.

In this framework, we address the following challenges. First, coexistence fairness should be ensured. Unlike classic monetary spectrum trading that leads to prioritized spectrum access, our framework aims to ensure fair spectrum sharing, since all offerers and requesters are equal secondary networks. In auction theory, social welfare characterizes the efficiency of resource reallocation, and is an indication of coexistence fairness. Maximizing social welfare avoids the unfair case where under-utilized spectrum is reallocated to a network who has less interest in it. To achieve this, our auction design has to thwart untruthful reporting of spectrum valuations from bidding requesters. Second, spectrum utilization should be maximized dynamically. When the availability of auctioned spectrum is very limited, our auction design has to consider spatial reuse of the spectrum. In a case where bidding requesters can dynamically arrive and depart, online updating of spectrum sharing results has to be supported. In addition, spectrum utilization needs to be maximized to improve each offerer's revenue in a multi-winner auction. Third, the convergence of our framework to a near-optimal equilibrium solution should be guaranteed in a distributed

manner. Without a centralized coexistence infrastructure, such as the IEEE 802.19.1 system [15], each requester in a non-cooperative game is likely to act greedily. In the presence of conflicts of interest, our game and auction designs have to always guarantee system stability.

To address the above challenges, the contributions of this chapter are summarized as follows.

- We propose a game-auction coexistence framework, which incorporates a game problem for supporting distributed spectrum sharing and an auction problem for ensuring system stability, coexistence fairness, and spectrum reuse.
- We address the trade-offs among social welfare, offerer's revenue, and requester's utility. The offerer's strategy tries to maximize social welfare as well as offerer's revenue in the auction, while the requester's strategy aims to maximize requester's utility in the game. The proposed strategies guarantee truthful, fair, and efficient spectrum reallocation.
- We prove that the proposed coexistence framework is stable.
- Our simulation results show that the proposed framework always converges to a near-optimal distributed solution and improves coexistence fairness and spectrum utilization.

The remainder of this chapter is organized as follows. Related work is discussed in section 5.2. The game-auction coexistence framework is introduced in section 5.3. The problems of auction and game are studied in sections 5.4 and 5.5, respectively. System stability is proved in section 5.6. Implementation issues are discussed in section 5.7. Simulation results are presented in section 5.8. The chapter is summarized in section 5.9.

5.2 Related Work

In IEEE 802.16h, CT-CXP enables dynamic subframe sharing among co-located coexisting systems [11]. This protocol is based on the idea of auction-based spectrum leasing to reallocate under-utilized spectrum from offering networks to requesting networks. Each network

follows certain coexistence etiquette. However, CT-CXP only defines a protocol for the exchange of control messages between offerers and requesters, and it does not provide any details regarding actual algorithm designs. We study auction-based spectrum sharing in the context of a non-cooperative game, and consider many important coexistence issues, such as the trade-offs among objectives of offerers and requesters, spectrum reuse, online auction winning, bidding truthfulness, and system stability.

In CR networks, auction models are usually used to study centralized spectrum trading between primary spectrum sellers (or spectrum brokers) and secondary spectrum buyers [72, 73, 74, 75, 76, 77, 78, 79]. The maximization of seller's revenue is the primary objective in such monetary auction problems. In our spectrum etiquette framework, however, each offerer maximizes social welfare as well as its revenue to achieve equal spectrum sharing. The need for bidding truthfulness is important to prevent market manipulation [72, 73, 74, 75]. A truthful double auction is studied in [72], which requires a centralized entity. In contrast, we study a distributed coexistence problem. In [73], bidding truthfulness is achieved by utilizing the distribution of buyers' valuations, which is hard to get in practice. We do not assume such a priori knowledge. The truthful second-price auction is applied to a multi-winner case in [74], where a set of bidders is grouped as a "virtual bidder". However, each individual bidder can still untruthfully get extra benefits and harm other bidders in the same winning group. On the contrary, we ensure that each individual bidder stays truthful to its spectrum valuation. The idea of pricing strategy in [75] is to charge each winner by a critical value, but this method cannot be applied to a multi-winner auction, which is necessary to enable spatial reuse of auctioned spectrum. Bidding truthfulness is not addressed in [76, 77, 78, 79], which assume that each bidder always reports its true spectrum valuation. However, untruthful bidding can greatly harm the maximization of social welfare.

As for distributed spectrum trading between multiple sellers and multiple buyers, game models are used to characterize the competition among sellers who want to attract buyers so as to maximize their revenue [80, 81, 82, 83]. In the secondary spectrum market [83], for example, each secondary service provider buys spectrum from primary users and sells

the spectrum to secondary users to make a profit. The spectrum resource for sale directly comes from primary users and is thus relatively abundant. In contrast, we study the competition among requesters in a over-demand (instead of over-supply) case, where the spectrum resource from each secondary offerer is temporary and limited.

5.3 System Model

In this section, we discuss the system model that underpins the proposed CTSE, which is a game-auction coexistence framework.

5.3.1 Credit-Token-Based Spectrum Etiquette

The credit-token-based spectrum etiquette concept in IEEE 802.16h CT-CXP [11] serves as the basis of our auction problem. In this reservation-based coexistence protocol, each co-located network can be either an *offerer* or a *requester* to "sell" or "buy" the rights to access shared spectrum, respectively, in certain under-utilized time subframes for a finite amount of time. An auction is needed if an offerer's claimed subframes cannot accommodate all the requesters. In this case, the selected auction winners need to "pay" *credit tokens* for the granted spectrum access. A credit token is defined as a pseudo monetary unit. For each network, the amount of credit tokens it owns is proportional to its likelihood of getting fallow spectrum from other networks. The coexisting networks can exchange control messages locally by means of a common radio access technology and a common control channel. A mechanism like CT-CXP based on frame reservation has to rely on tight synchronization among coexisting networks, which is challenging to achieve when the coexisting heterogeneous networks use different time frame formats. Hence, we consider a coexistence framework based on channel reservation.

5.3.2 A Game-Auction Coexistence Framework

We study auction-based spectrum sharing in the context of a non-cooperative game, and consider many important coexistence issues, such as optimization trade-offs, spectrum reuse, online auction winning, bidding truthfulness, and system stability. As in Figure 5.1, we model the problem for CTSE as a *game-auction* coexistence framework.

Suppose that there are N co-located CR networks. In each network n for $n \in \mathcal{N} \triangleq \{1, \cdots, N\}$, the base station on behalf of the entire network claims a set of master channels, denoted by $\mathcal{M}_n$, via local channel reservation. The initial channel reservation can be on a first-come-first-serve basis, which is assumed in CT-CXP. The channels in $\mathcal{M}_n$ are fallow white spaces locally available to network n. Each network queries a geolocation database or performs spectrum sensing to identify the set of white-space channels and claims $\mathcal{M}_n$ from it. The number of master channels and the MAC protocol (with or without channel aggregation [10]) used to access those channels may be different for each network, since coexisting networks may use heterogeneous air interface standards. Other neighboring networks can still access $\mathcal{M}_n$ as slave channels under the control of network n. When a network o has surplus spectrum, it can act as an offerer and rent out a subset of $\mathcal{M}_o$, say $\mathcal{G}_o \subseteq \mathcal{M}_o$. The set of offerers is denoted by $\mathcal{O} \subseteq \mathcal{N}$. When a network r needs extra spectrum, it can act as a requester and bid for the access to $\mathcal{G}_o$ for $o \in \mathcal{O}$ using credit tokens. The set of requesters is denoted by $\mathcal{R} \subseteq \mathcal{N}$. We assume that co-located coexisting networks exchange control messages over a common control channel [84]. The initial channel reservation and fallow spectrum reallocation are limited to a local region. More details will be given in section 5.7. Unlike the fixed sellers and buyers in most existing work, the sets $\mathcal{O}$ and $\mathcal{R}$ dynamically change over time in our coexistence framework.

Upon receiving the announcement of $\mathcal{G}_o$ from offerer o, each requester $r \in \mathcal{R}$ first makes sure that $\mathcal{G}_o$ is locally available, and then comes up with a *valuation* of $\mathcal{G}_o$, denoted by v_{or}. Different networks may have different valuations for the same spectrum according to their special needs. Among the set of available offerers, denoted by $\mathcal{O}_r \subseteq \mathcal{O}$, who have

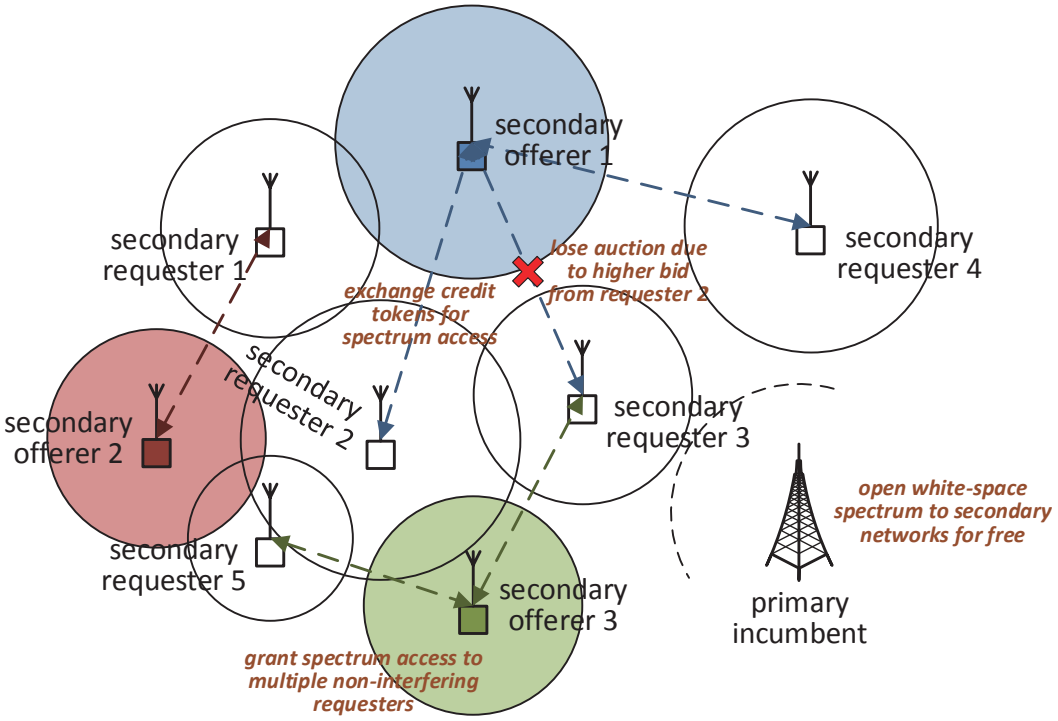

Figure 5.1: An example of credit-token-based spectrum etiquette.

announced fallow spectrum locally available to requester r, a set of preferred offerers, denoted by $\mathcal{D}_r \subseteq \mathcal{O}_r$, are selected for bidding. In section 5.5, we will discuss the requester's dynamic offerer selection strategy that maximizes *requester's utility*, denoted by U_r. For each preferred offerer, say o, requester r submits a *bid*, denoted by b_{or}, showing interest for the channels in $\mathcal{G}_o$. Note that the value of b_{or} is not necessarily equal to v_{or}. We assume that the amount of credit tokens owned by each network is limited to avoid arbitrarily large bids over valuations. Unlike the unevenly distributed monetary credits for prioritized spectrum access, credit tokens can be carefully regenerated when needed for fair spectrum sharing, e.g., when new users become active. Each requester can be greedy in maximizing its utility regardless of interfering others. Hence, we frame the offerer selection problem as a non-cooperative game, in which each requester acts as a player.

As for each offerer $o \in \mathcal{O}$, an auction is held periodically until the offerer becomes a requester when short of spectrum. Each auction period consists of an initial auction phase and an

online auction phase. The durations of these phases are offerer-dependent, which can be determined by channel availability, traffic load, and control overhead. For example, if the spectrum offered by a database-driven white space network is found to be locally available for a long time, the auction periods can be relatively long for minimal overhead. In the initial auction phase, offerer o maintains a set of interested requesters, denoted by $\mathcal{R}_o \subseteq \mathcal{R}$. At this point, offerer o has collected a set of bids, denoted by $\mathcal{B}_o$, which have been submitted by the requesters in $\mathcal{R}_o$ according to the set of valuations, denoted by $\mathcal{V}_o$. Given $\mathcal{B}_o$, offerer o selects a set of *winners*, denoted by $\mathcal{W}_o \subseteq \mathcal{R}_o$, to grant them the access to $\mathcal{G}_o$. In section 5.4, we will discuss the offerer's dynamic winner selection strategy that maximizes *social welfare*, denoted by S_o. The set of *losers* is denoted by $\mathcal{L}_o \triangleq \mathcal{R}_o - \mathcal{W}_o$. Due to spectrum reuse, multiple winners can share the same spectrum at different locations. In the online auction phase, interested requesters can dynamically arrive and depart between two successive initial auction phases. Hence, each $\mathcal{W}_o$ is subject to changes anytime in such an online multi-winner auction. For each winner, say $w \in \mathcal{W}_o$, offerer o computes a *price*, denoted by p_{ow}, as the amount of credit tokens that winner w needs to pay in a certain auction period. The set of prices is denoted by $\mathcal{P}_o$. Because bidding truthfulness is important for the maximization of social welfare, we study the spectrum sharing problem at each offerer as a truthful multi-winner auction. Besides, the maximization of *offerer's revenue*, denoted by Q_o, is also addressed given guaranteed system stability.

In summary, the above problem for CTSE is able to be formulated as a hybrid game-auction coexistence framework. The requesters compete with each other in a non-cooperative game, and the offerers hold truthful multi-winner auctions to allocate their surplus spectrum resource to bidding requesters. The interactions between the requesters' offerer selection and bidding strategies and the offerers' winner selection and pricing strategies determine the stability, fairness, and efficiency of our coexistence framework.

5.4 Auction Problem

In this section, we focus on the auction problem for a certain offerer, say $o \in \mathcal{O}$. The multi-offerer scenario will be studied in the context of a game. We propose the offerer's strategy that maximizes social welfare and guarantees bidding truthfulness.

5.4.1 Initial and Online Auctions

In each initial auction phase, offerer o selects winners $\mathcal{W}_o$ from interested requesters $\mathcal{R}_o$ to enable spectrum sharing among the selected winners. Here, $\mathcal{R}_o$ is given by the game problem in our game-auction framework. We assume that there is one channel available in $\mathcal{G}_o$, which can be shared by all the requesters in $\mathcal{R}_o$. Note that the formulation can be readily extended to a multi-channel case. The primary objective of credit-token-based auction problem is to maximize social welfare, which characterizes the efficiency of spectrum reallocation, i.e., coexistence fairness. Under-utilized spectrum needs to be reallocated to the networks who show the greatest interest in it, since all offerers and requesters are equal to share white-space spectrum for free. Hence, we borrow the concept of credit tokens from CT-CXP to enable fair spectrum sharing. But in section 5.6, we do consider the secondary objective to maximize offerer's revenue, which is more important in classic monetary auction problems. The social welfare achieved by offerer o, i.e., S_o, can be defined by

$$S_o \triangleq \sum_{r \in \mathcal{R}_o} x_{or} v_{or}, \tag{5.1}$$

where x_{or} is a binary indicator such that $x_{or} = 1$ (0) represents requester r is (is not) selected as a winner. The S_o is the sum of winners' valuations of $\mathcal{G}_o$. In a social welfare maximization (SWM) problem, offerer o finds a $\mathcal{W}_o$ that maximizes S_o.

Because all the winners in $\mathcal{W}_o$ share the same $\mathcal{G}_o$ at the same time, offerer o has to ensure that the winners do not interfere with each other. We assume that an interference graph, denoted

by $\mathbf{I}_o \triangleq \{a_{ij}\}_{i,j\in\mathcal{R}_o}$, is generated at offerer o, where a_{ij} is a binary indicator such that $a_{ij} = 1$ (0) indicates requesters i and j are (are not) interfering neighbors. The interference graph can be generated by integrating neighbor sets, which are reported from interested requesters along with their bids [72, 73, 74, 76]. Two neighbors are viewed to be interfering when at least one of them is interfered. Interfering neighbors cannot win together at the same offerer in the same auction period. Hence, the constraint for winner selection is expressed by

$$a_{ij}(x_{oi} + x_{oj}) \leq 1, \text{ for } i, j \in \mathcal{R}_o. \tag{5.2}$$

We summarize the SWM problem at each offerer o as follows.

Problem 1 (SWM)

Find: x_{or} for all $r \in \mathcal{R}_o$;

Maximize: S_o;

Subject to: (5.2).

Given $\mathcal{R}_o$ and $\mathcal{V}_o$ from the game problem, Problem 1 is similar to the binary integer programming problem in [74], which was shown to be solvable. However, solving Problem 1 requires offerer o to know the true valuations $\mathcal{V}_o$, whereas offerer o only has a knowledge of the collected bids $\mathcal{B}_o$. Hence, bidding truthfulness, i.e., $\mathcal{B}_o = \mathcal{V}_o$, needs to be guaranteed.

In each online auction phase, the offerer's dynamic winner selection strategy also supports online updating of $\mathbf{I}_o$ and $\mathcal{W}_o$, which is initially given by solving Problem 1. Unlike a classic auction that always restricts winner selection at a specific time, the auction at offerer o allows online bidding and winning at any time. During this auction phase, there are three possible arrival or departure events at offerer o:

a. A requester arrives and does not cause harmful interference to any existing winner in $\mathcal{W}_o$;

b. A requester arrives and causes harmful interference to at least one existing winner in $\mathcal{W}_o$;

c. A loser in $\mathcal{L}_o$ departs.

The corresponding offerer's strategy is proposed as follows. For case a, the arrived requester, say w, can join in $\mathcal{W}_o$ directly to further increase S_o to $S_o + v_{ow}$ without waiting for the next

auction period. For case b, the arrived requester cannot be added to $\mathcal{W}_o$, since it harms S_o due to the drop of v_{ow} for any interfered winner $w \in \mathcal{W}_o$. For case c, nothing needs to be done, since S_o is still maximized by the unchanged $\mathcal{W}_o$. We ignore the case that a winner in $\mathcal{W}_o$ departs, which decreases both social welfare and the winner's utility.

5.4.2 Bidding Truthfulness

The offerer's pricing strategy is another key issue in an auction problem. In the presence of untruthful bidding, social welfare cannot be maximized. Intuitively, two neighboring requesters, say i and j, submit untruthful bids $b_{oi} < v_{oi}$ and $b_{oj} < v_{oj}$ to offerer o, respectively, and $b_{oi} > b_{oj}$ while $v_{oi} < v_{oj}$. But requester i is the winner even though requester j needs the spectrum more, i.e., $v_{oi} < v_{oj}$.

A common way of guaranteeing bidding truthfulness, i.e., $\mathcal{B}_o = \mathcal{V}_o$, is to charge each winner w by a price $0 \leq p_{ow} < b_{ow}$ and p_{ow} is not a function of b_{ow} so as to force requesters to report true valuations. The classic Vickrey-Clarke-Groves (VCG) auction [85] belongs to this category and is appropriate for our coexistence problem. The objective of a VCG auction is to maximize social welfare truthfully, which matches with our objective in (5.1). The major drawback of a VCG strategy is that seller's revenue can be very low (even zero) in some cases [74]. However, in our coexistence framework, spectrum reuse allows each offerer to collect considerable credit tokens from multiple winners, even without giving up the offered spectrum (by treating itself as a winner). This will be shown in section 5.8. Furthermore, some extra credit tokens may be generated as a bonus to compensate offerer's loss of revenue when needed, which will be discussed in section 5.7.

Based on the VCG pricing strategy, the price (credit tokens per unit time) for each winner w, i.e., p_{ow}, can be defined by

$$\begin{aligned} p_{ow} &\triangleq b_{ow} - \Big(\sum_{w' \in \mathcal{W}_o} b_{ow'} - f_w(\mathbf{b}_{-w}) \Big) \\ &= f_w(\mathbf{b}_{-w}) - \sum_{w' \in \mathcal{W}_o - \{w\}} b_{ow'}, \text{ for } w \in \mathcal{W}_o, \end{aligned} \tag{5.3}$$

where $\mathbf{b}_{-w} = \mathcal{B}_o - \{b_{ow}\}$, and f_w is a certain pricing function in general form defined for winner w. Here, f_w needs to satisfy

$$\sum_{w' \in \mathcal{W}_o} b_{ow'} - f_w(\mathbf{b}_{-w}) > 0, \tag{5.4}$$

which is used to guarantee $p_{ow} < b_{ow}$, and

$$\sum_{w' \in \mathcal{W}_o} b_{ow'} - f_w(\mathbf{b}_{-w}) \leq b_{ow}, \tag{5.5}$$

which is used to guarantee $p_{ow} \geq 0$. The choice of f_w does not affect bidding truthfulness but does determine system stability, which will be discussed in section 5.6. For each loser, say $l \in \mathcal{L}_o$, we assume that $p_{ol} = 0$. Then, the revenue achieved by offerer o, i.e., Q_o, through the auction can be defined by

$$Q_o \triangleq \sum_{w \in \mathcal{W}_o} p_{ow}. \tag{5.6}$$

In section 5.6, offerer's revenue Q_o can be maximized subject to system stability, even though social welfare S_o is the primary objective of offerer o.

In each online auction phase, $\mathcal{W}_o$ is updated by case a as discussed above. Hence, the offerer's pricing strategy needs to consider such dynamic change of $\mathcal{W}_o$. For case a, $\mathcal{R}_o$ and $\mathcal{W}_o$ are added with the newly joined winner, say w'. But its price $p_{ow'}$ can still be determined by (5.3) given a certain definition of $f_{w'}$. The price for an existing winner, say $w \in \mathcal{W}_o$, does not have to be changed for all the three cases.

Lemma 1 The pricing strategy of each offerer o for $o \in \mathcal{O}$ defined by (5.3) guarantees bidding truthfulness, i.e., $\mathcal{B}_o = \mathcal{V}_o$, regardless of f_w for $w \in \mathcal{W}_o$.

Proof For each winner w, its auction benefit is determined by $v_{ow} - p_{ow}$, which is not a function of its bid b_{ow}. Hence, bidding with $b_{ow} = v_{ow}$ is always the best strategy for each requester so as to avoid negative outcomes, either higher chance to lose due to $b_{ow} < v_{ow}$ or higher chance to pay $p_{ow} > v_{ow}$ due to $b_{ow} > v_{ow}$. □

5.5 Game Problem

In this section, we model the game problem for an entire system, in which multiple offerers and multiple requesters coexist. We propose the requester's strategy accordingly.

In the non-cooperative game where each requester acts as a player, if two requesters are interfering neighbors and bid at the same offerer, at most one of them can obtain positive utility. Hence, there exist conflicts of interest among interfering requesters. Each requester r selects preferred offerers $\mathcal{D}_r$ from $\mathcal{O}_r$ that maximizes utility, i.e., U_r, which is defined by

$$U_r \triangleq \sum_{o \in \mathcal{O}_r} y_{or} U_{or} = \sum_{o \in \mathcal{O}_r} y_{or}(x_{or} v_{or} - p_{or}), \tag{5.7}$$

where y_{or} is a binary indicator such that $y_{or} = 1$ (0) represents requester r chooses (does not choose) offerer o, and $U_{or} \triangleq x_{or} v_{or} - p_{or}$. In a requester's utility maximization (RUM) problem, requester r wants to find $\mathcal{D}_r$ who provide the most valuable spectrum and charge the lowest prices in order to maximize U_r. To clearly study the move actions of requester r in the game, we assume that requester r selects at most one offerer at a time. The constraint for offerer selection is expressed by

$$\sum_{o \in \mathcal{O}_r} y_{or} \leq 1. \tag{5.8}$$

We are able to write the RUM problem at each requester r as follows.

Problem 2 (RUM)

Find: y_{or} for all $o \in \mathcal{O}_r$;

Maximize: U_r;

Subject to: (5.8).

Given the values of x_{or} and p_{or} from the auction problem at offerer o for $o \in \mathcal{O}_r$, Problem 2 can be solved directly via brute force, since $x_{or}v_{or}$ and p_{or} defining U_r are viewed as constants. But, solving Problem 2 requires requester r to know auction results from other offerers, which is hard to implement in practice. The related issues will be discussed in section 5.7.

5.6 System Stability

In the game problem, each requester solving Problem 2 may switch from one preferred offerer to another. This can in return trigger the affected offerer solving Problem 1 to change the sets of winners and prices in the auction problem. The changes of winners and prices may again trigger more requesters to switch between different offerers. Hence, such chain reactions can result in an unstable system. In this section, we describe the offerer's pricing strategy that guarantees system stability.

Each offerer o is able to observe a population of interested requesters $\mathcal{R}_o$. The size of each population dynamically changes as requesters switch between different offerers. According to the Lyapunov's direct stability theorem in nonlinear control theory, if there exists a *Lyapunov function* on a region containing a *critical point*, which is set as the origin, then the origin is a stable critical point and all solutions originating in the region approach the origin asymptotically [71]. Hence, if we can find a critical point and a Lyapunov function on a region of the state space $\mathbf{R}$, where a state is defined by $\mathbf{r} = \{\mathcal{R}_o \mid o \in \mathcal{O}\} \in \mathbf{R}$, the coexistence system is guaranteed to be stable.

The gain of utility is the incentive of each requester to take a switching action, i.e., a change of preferred offerer. If any requester cannot find a better offerer, the system converges to an

equilibrium point, a critical point. The total utility of requesters in $\mathcal{R}$, denoted by U^L, is

$$U^L \triangleq \sum_{o\in\mathcal{O}} U_o^L = \sum_{r\in\mathcal{R}} U_r, \tag{5.9}$$

where $U_o^L \triangleq \sum_{r\in\mathcal{R}_o} U_r = \sum_{r\in\mathcal{R}_o} U_{or}$. The U^L characterizes the incentive of requesters to take switching actions. A higher value of U^L means less incentive of requesters. We show that $\hat{U}^L - U^L$ is a Lyapunov function under a carefully designed pricing function f_w, where $\hat{U}^L$ is the upper bound of U^L.

In (5.7), U_r depends on $x_{or}v_{or} - p_{or}$ for $o \in \mathcal{O}_r$. Know that $p_{ol} = 0$ for $l \in \mathcal{L}_o$, while p_{ow} for $w \in \mathcal{W}_o$ depends on the choice of f_w as needed in (5.3). Each offerer o can define a pool of f_w for $w \in \mathcal{W}_o$ to control the value of U_o^L. For a winner w, one possible definition of f_w is

$$f_w(\mathbf{b}_{-w}) \triangleq \alpha_w \sum_{w'\in\mathcal{W}_o^{-w}} v_{ow'}, \tag{5.10}$$

where $\mathcal{W}_o^{-w}$ is the set of winners obtained by solving Problem 1 given that winner w is not present, and $\alpha_w \in (0, 1]$ is a control factor. The physical meaning of p_{ow} given by (5.3) and (5.10) is the harm caused by winner w's presence to other requesters in $\mathcal{R}_o$. In the rest of this section, we show how this definition of f_w for every winner w ensures system stability.

We now focus on an offerer o to study the change of U_o^L as a requester arrives or departs. In each initial auction phase, there are three possible cases for the arrival or departure of a requester at offerer o solving Problem 1:

1. A requester w' arrives and becomes a winner without harming the existing winners in $\mathcal{W}_o$ (original);
2. A requester w' arrives and becomes a winner by forcing a set of existing winners, $\mathcal{W}^l \subseteq \mathcal{W}_o$ (original), to be losers;
3. An existing loser l in $\mathcal{L}_o$ departs.

Note that the three cases described here affect the solution to Problem 1, while the three cases listed in section 5.4 are used to support extra online auction after solving Problem 1.

We will discuss the impact of online auction on system stability later. According to (5.7), we know that $U_{ow} = v_{ow} - p_{ow} > 0$ for $w \in \mathcal{W}_o$, while $U_{ol} = 0$ for $l \in \mathcal{L}_o$. Hence, a loser always has greater incentive to take a switching action than a winner does. We assume that each requester always bids for the channel with the currently highest valuation. Once the requester wins, it can stick to this channel if no new resource becomes available within this auction period, since the channel is mostly preferred and the requester's contribution to social welfare is maximized. Hence, we ignore the counterintuitive case that an existing winner departs. We will show that social welfare and requester' utility are maximized under such a strategy. If a loser cannot win a channel when its bid is much lower than the winning bids, it can quit the system, since its utility is always zero and cannot be improved.

Now, we prove system stability formally. Several Lemmas are presented to use the Lyapunov's direct stability theorem.

Lemma 2 At each offerer o for $o \in \mathcal{O}$, for case 1 and case 3, $U_o^L = \sum_{w \in \mathcal{W}_o} U_{ow}$ is increasing. For case 2, U_o^L is increasing when $\alpha_{w'}$ in (5.10) satisfies

$$\check{\alpha}_{w'} \leq \alpha_{w'} < \hat{\alpha}_{w'},$$

where $\check{\alpha}_{w'}$ is a lower bound making $p_{ow'} = 0$, and $\hat{\alpha}_{w'}$ is an upper bound given by $\mathcal{V}_o$ and α_w for $w \in \mathcal{W}_o - \mathcal{W}^l$.

Proof For both case 1 and case 3, we consider the worst case where $\alpha_{w'} = 1$. For case 1, $U_{ow'} > 0$ and the values of U_{ow} are non-decreasing for $w \in \mathcal{W}_o$. For case 3, if $l \in \mathcal{W}_o^{-w}$ for a certain $w \in \mathcal{W}_o$, the value of p_{ow} is decreased due to the decreased $f_w(\mathbf{b}_{-w})$. Then, the value of U_{ow} is increased. If $l \notin \mathcal{W}_o^{-w}$ for a certain $w \in \mathcal{W}_o$, the value of U_{ow} is unchanged. Hence, U_o^L is increasing for case 1 and case 3.

For case 2, we assume the worst case that all the new losers in $\mathcal{W}^l$ drop in $\mathcal{W}_o^{-w}$ and no existing requesters in $\mathcal{W}_o^{-w}$ are removed for $w \in \mathcal{W}_o - \mathcal{W}^l$ (including w'). In (5.10), the maximized size of $\mathcal{W}_o^{-w}$ causes the maximal possible drop of U_o^L. For the new winner w',

the utility gain $\Delta U_{ow'} = (\sum_{w\in\mathcal{W}_o-\mathcal{W}^l} v_{ow} + v_{ow'}) - \alpha_{w'}(\sum_{w\in\mathcal{W}_o-\mathcal{W}^l} v_{ow} + \sum_{l\in\mathcal{W}^l} v_{ol})$. For each remaining winner $w \in \mathcal{W}_o - \mathcal{W}^l$, the utility gain $\Delta U_{ow} = v_{ow'} - (1+\alpha_w)\sum_{l\in\mathcal{W}^l} v_{ol}$. For each new loser $l \in \mathcal{W}^l$, the utility gain $\Delta U_{ol} = -(\sum_{w\in\mathcal{W}_o-\mathcal{W}^l} v_{ow} + \sum_{l\in\mathcal{W}^l} v_{ol}) + f_l(\mathbf{b}_{-l})$. Hence, to guarantee that U_o^L is always increasing, we make the total utility gain $\Delta U_o^L > 0$, namely

$$\Delta U_o^L = \Delta U_{ow'} + \sum_{w\in\mathcal{W}_o-\mathcal{W}^l} \Delta U_{ow} + \sum_{l\in\mathcal{W}^l} \Delta U_{ol} > 0. \tag{5.11}$$

We can control $\alpha_{w'}$ (and α_w for $w \in \mathcal{W}_o - \mathcal{W}^l$) to make

$$\Delta U_o^L \triangleq (\hat{\alpha}_{w'} - \alpha_{w'})(\sum_{w\in\mathcal{W}_o-\mathcal{W}^l} v_{ow} + \sum_{l\in\mathcal{W}^l} v_{ol}) > 0,$$

where $\hat{\alpha}_{w'}$ is a function of $\mathcal{V}_o$ and α_w for $w \in \mathcal{W}_o - \mathcal{W}^l$ based on (5.11). Hence, we have $\alpha_{w'} < \hat{\alpha}_{w'}$. Note that such an $\alpha_{w'}$ always exists. Even for the worst case that U_o^L is already maximized by $p_{ow} = 0$ for $w \in \mathcal{W}_o$ (original) before the arrival of winner w', a solution to make $\Delta U_o^L > 0$ is $\alpha_w = \check{\alpha}_w$ making $p_{ow} = 0$ for $w \in \mathcal{W}_o - \mathcal{W}^l$ (including w'), since $v_{ow'} > \sum_{l\in\mathcal{W}^l} v_{ol}$. Hence, U_o^L can be increasing for case 2. □

Furthermore, we are able to show that ΔU_o^L is non-increasing after certain actions of requesters in the game.

Lemma 3 At each offerer o for $o \in \mathcal{O}$, after $p_{ow} = 0$ for all $w \in \mathcal{W}_o$ and $\mathcal{W}_o$ is unchanged, $U_o^L = \sum_{w\in\mathcal{W}_o} U_{ow}$ reaches its upper bound, say $\hat{U}_o^L$, i.e., $\Delta U_o^L = 0$.

Proof A rational requester always selects an offerer who achieves the currently highest valuation, so each winner w at offerer o maximizes $x_{ow}v_{ow}$. If winner w can also minimize p_{ow} at the same time, then U_{ow} is maximized. Fortunately, after certain switching actions of requesters in $\mathcal{R}$, we can have $p_{ow} = 0$ for $w \in \mathcal{W}_o$. As more losers quit the coexistence system due to too low bids and valuations, there is higher chance that $\mathcal{W}_o^{-w} = \mathcal{W}_o - \{w\}$ occurs. For each winner w, when $\mathcal{W}_o^{-w} = \mathcal{W}_o - \{w\}$, we have $p_{ow} = 0$ according to (5.3) and (5.10) for $\alpha_w = 1$. In the end, all the remaining requesters in the system are winners,

and all the losers have quitted due to always zero utility. Once $\mathcal{W}_o$ keeps unchanged, i.e., social welfare is maximized at offerer o, and $p_{ow} = 0$ for all $w \in \mathcal{W}_o$, U_o^L is maximized to $\hat{U}_o^L \triangleq \sum_{w \in \mathcal{W}_o} v_{ow}$, and thus $\Delta U_o^L = 0$. □

Then, we study the change of U_o^L as a requester arrives or departs in an online auction phase.

Lemma 4 At each offerer o for $o \in \mathcal{O}$, online auction does not decrease U_o^L, and still guarantees $\Delta U_o^L = 0$.

Proof In the three cases described in section 5.4, case a increases U_o^L, which can be proved as case 1 in Lemma 2. For case b and case c, $\mathcal{W}_o$ is not changed. The prices for existing winners in $\mathcal{W}_o$ do not have to be changed. Then, U_o^L is not changed by case b and case c. Hence, online auction does not decrease U_o^L. After all the losers quit the coexistence system due to zero utility, we still have $\Delta U_o^L = 0$. □

Because a higher value of U^L indicates less incentive of requesters to change their preferred offerers, we can expect that the coexistence system gradually approaches and finally reaches an equilibrium point. We now prove this conjecture formally using the Lyapunov's direct stability theorem.

Theorem 1 For certain $\mathcal{O}$ and $\mathcal{R}$, the coexistence system is defined by Problem 1 at each offerer o for $o \in \mathcal{O}$ and Problem 2 at each requester r for $r \in \mathcal{R}$. The offerer's pricing strategy is defined by (5.3) and (5.10). Given an initial point, say $\mathbf{r}_0 = \{\mathcal{R}_o \mid o \in \mathcal{O}\} \in \mathbf{R}$, when (5.11) holds, the coexistence system converges to a stable equilibrium point, say $\mathbf{r}^* \in \mathbf{R}$, where U^L is maximized and $\Delta U_o^L = 0$ for all $o \in \mathcal{O}$.

Proof First, we show that $\mathbf{r}^*$ is a critical point. Because $\Delta U_o^L = 0$ for all $o \in \mathcal{O}$ at $\mathbf{r}^*$, then $\Delta U^L \triangleq \sum_{o \in \mathcal{O}} \Delta U_o^L = 0$ at $\mathbf{r}^*$, which accords with the definition of critical point. We set $\mathbf{r}^*$ as the origin of a region.

Next, we prove that $\hat{U}^L - U^L$ is a Lyapunov function, where $\hat{U}^L \triangleq \sum_{o \in \mathcal{O}} \hat{U}_o^L$. If $\hat{U}^L - U^L$ is indeed a Lyapunov function, it needs to satisfy the following conditions [71]:

1) To show $\hat{U}^L - U^L = 0$ at the origin, i.e., $\mathbf{r}^*$, and $\hat{U}^L - U^L > 0$ for other $\mathbf{r} \in \mathbf{R}$;

2) To show $-\frac{dU^L}{d\mathbf{p}} > 0$, where $\mathbf{p}$ is any vector radiating from the origin, i.e., $\mathbf{r}^*$;

3) To show $-\frac{dU^L}{dt} < 0$, where t is the time.

According to Lemma 3, U^L can reach its upper bound $\hat{U}^L$. Clearly, $\hat{U}^L - U^L = 0$ at $\mathbf{r}^*$ and $\hat{U}^L - U^L > 0$ for other $\mathbf{r} \in \mathbf{R}$. Hence, condition 1) is satisfied. According to Lemma 2, U^L is increasing in time. Hence, condition 3) is satisfied. We know that $\mathbf{r}^*$ is the highest point in terms of U^L, so U^L is decreasing via $\mathbf{p}$ radiating from $\mathbf{r}^*$. Hence, condition 2) is satisfied. Then, we can say that $\hat{U}^L - U^L$ is a Lyapunov function.

According to the Lyapunov's direct stability theorem, we conclude that $\mathbf{r}^*$ is a stable critical point and the coexistence system starting from any $\mathbf{r}_0$ approaches $\mathbf{r}^*$ asymptotically. □

We now relax the definition of f_w in (5.10) for the offerer's pricing strategy to a general setting. In this case, each offerer o can dynamically select a f_w for each winner w from a pool of definitions to maximize its revenue, i.e., Q_o, given guaranteed system stability by Theorem 1. Besides Problem 1, each offerer o can solve an offerer's revenue maximization (ORM) problem, which is written as follows.

Problem 3 (ORM)

$$\begin{aligned} \text{Find:} \quad & f_w \text{ for all } w \in \mathcal{W}_o; \\ \text{Maximize:} \quad & Q_o; \\ \text{Subject to:} \quad & (5.4), (5.5), (5.11). \end{aligned}$$

Therefore, our game-auction coexistence framework addresses the trade-offs among social welfare and offerer's revenue in the auction and requester's utility in the game.

Corollary 1 The coexistence system that is defined by Problem 1 and Problem 3 at each offerer and Problem 2 at each requester is guaranteed to be stable.

Proof At each offerer, the choice of pricing functions is constrained by (5.4), (5.5), and (5.11). In the general setting, Lemma 2 still hold due to (5.11), and Lemma 3 still hold due to increasing and bounded total utility (bounded total price) by (5.4) and (5.5). Hence, the coexistence system defined in Theorem 1 is stable. Because solving Problem 3 at each offerer

Algorithm 3 Offerer o's strategy for $o \in \mathcal{O}$

1: **check** the availability of surplus spectrum $\mathcal{G}_o$ for the current auction period (via geolocation database or spectrum sensing), and broadcast it to requesters $\mathcal{R}$
2: **collect** bids $\mathcal{B}_o$ (i.e., valuations $\mathcal{V}_o$ of $\mathcal{G}_o$) from interested requesters $\mathcal{R}_o$, and update interference graph $\mathbf{I}_o$
3: **solve** Problem 1 to select a set of non-interfering winners $\mathcal{W}_o$ according to $\mathcal{B}_o$ and $\mathbf{I}_o$
4: **solve** Problem 3 to select a pricing function f_w for each winner w for $w \in \mathcal{W}_o$
5: **compute** a price p_{ow} for each winner w according to (5.3), and obtain prices $\mathcal{P}_o$ and a bonus $\tilde{p}_o$
6: **allocate** $\mathcal{G}_o$ to each winner w, and broadcast winning/ losing notification along with $\mathcal{W}_o$ and $\mathcal{P}_o$
7: **while** in the online auction phase **do**
8: **collect** a bid from an online requester r
9: **if** case a in section 5.4 **then**
10: **allocate** $\mathcal{G}_o$ to requester r, charge a price p_{or}, and broadcast updated notification
11: **end if**
12: **end while**

does not change the winner selections at offerers and the offerer selections at requesters, system stability is still guaranteed. □

5.7 Algorithms and Implementation

In this section, we summarize the offerer's and requester's strategies in our coexistence framework, and discuss related implementation issues.

5.7.1 Offerer's Strategy

The proposed offerer's strategy as described in Algorithm 3 can be tailored to support the offering procedure defined in CT-CXP, including offering advertisement, renting request, and resource allocation. Specifically, line 1 implements offering advertisement. Line 2 enables renting request. Lines 3 to 6 realize not only the basic resource allocation required in CT-CXP but also spectrum reuse. Lines 7 to 12 define the extra online auction. The pricing in

Algorithm 4 Requester r's strategy for $r \in \mathcal{R}$

1: **collect** announcement messages about under-utilized spectrum from offerers $\mathcal{O}$
2: **check** the availability of offered spectrum for the current auction period locally (via geolocation database or spectrum sensing), and obtain available offerers $\mathcal{O}_r$
3: **solve** Problem 2 to select a set of preferred offerers $\mathcal{D}_r$
4: **submit** a bid $b_{or} = v_{or}$ to each offerer o for $o \in \mathcal{D}_r$, and wait for winning/losing notification (online requester)
5: **if** winning at offerer o **then**
6: **access** $\mathcal{G}_o$ directly
7: **else**
8: **try** other offerers, or wait for offerer o's next auction period (regular requester)
9: **end if**

lines 5 and 10 guarantees bidding truthfulness, which is neglected in CT-CXP. A bonus of credit tokens, denoted by $\tilde{p}_o$, can be added to each offerer o's revenue Q_o, which is used to encourage offerer o to participate in spectrum etiquette and maximize social welfare. As one possible definition of credit token bonus, $\tilde{p}_o$ can be proportional to the achieved social welfare S_o. In this case, the maximization of social welfare further contributes to the maximization of offerer's revenue. We plan to make the bonus design as part of our future work.

5.7.2 Requester's Strategy

The proposed requester's strategy as described in Algorithm 4 can support the requesting procedure defined in CT-CXP. Specifically, lines 1 to 2 correspond to offering advertisement. Lines 3 to 4 define renting request. A naive way of implementing line 3 would require each requester to know the exact value of achieved utility at each offerer to solve Problem 2, which is hard to achieve in practice. In a more effective approach, each requester can solve Problem 2 to maximize expected utility, which is computed based on a certain probability distribution. Alternatively, each requester can keep trying other offerers and switch to a new offerer only after it has won and has improved utility at the offerer. Lines 5 to 9 deal with resource allocation. The interactions between Algorithms 3 and 4 guarantee the stability of our coexistence framework, which is also neglected in CT-CXP.

5.7.3 Implementation Discussions

Because credit tokens can be flexibly generated or renewed for various purposes, e.g., fair spectrum sharing in our coexistence framework, the legitimacy of credit tokens in circulation should be ensured by a third-party entity. As guided by CT-CXP, coexisting networks can follow some spectrum etiquette to achieve efficient spectrum reallocation.

The CTSE framework relies on a common control channel [84] to support the exchange of control messages between co-located offerers and requesters. For heterogeneous CR networks with geolocation capability, Internet connection to a common geolocation database can be used for coordinated channel reservation [5]. For better real-time performance, e.g., online auction bidding and winning, each coexisting network may be equipped with an extra radio with common radio access technology for local exchange of control messages. A contention-based MAC protocol, such as CSMA, can be used to deal with control message collisions, especially in any offerer's initial auction phases. The message collisions can be mitigated by asynchronous offerers and prolonged auction periods. Each offerer can collect bids not only in the current initial auction phase but also in previous auction periods.

5.8 Performance Evaluation

In this section, we evaluate our coexistence framework by simulation. In order to clearly show the convergence process, we temporarily fix the number of offerers, i.e., $N_o \triangleq |\mathcal{O}|$, and the number of requesters, i.e., $N_r \triangleq |\mathcal{R}| = 10 \times N_o$. Any change of $\mathcal{O}$ or $\mathcal{R}$ results in a new coexistence scenario. Each spectrum valuation $v_{or} \in [0,1]$ for $o \in \mathcal{O}$ and $r \in \mathcal{R}$ and each interference indicator $a_{ij} \in \{0,1\}$ for $i,j \in \mathcal{R}$ are randomly generated. An interference graph $\mathbf{I}_o = \{a_{ij}\}_{i,j\in\mathcal{R}_o}$ is created at each offerer o according to the current $\mathcal{R}_o$. The value of v_{or} is set to zero when offerer o is not in $\mathcal{O}_r$. Each offerer runs Algorithm 3, and each requester runs Algorithm 4. The offerer's pricing strategy is defined by (5.3) and (5.10). To characterize the distributed nature of CR networks, we assume that interested requesters arrive at each offerer according to a Poisson process. The coexisting networks are asynchronous.

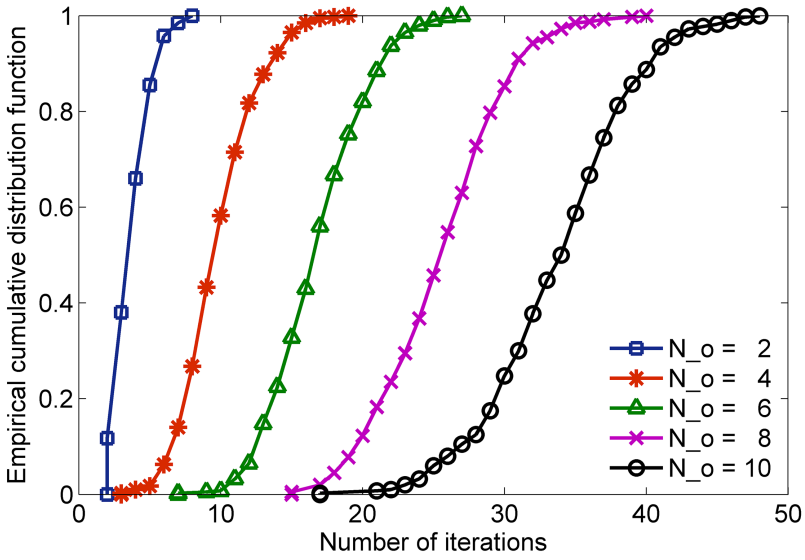

Figure 5.2: Convergence speed of CTSE framework.

First, we evaluate the convergence speed of proposed framework. We are interested in the number of iterations for a convergence process. Here, an "iteration" represents a switching action of any requester changing offerer selection in the game. The time duration of an iteration depends on arrival process, the durations of offerer-dependent auction periods, and the complexity of requester-dependent offerer selection. For certain fixed N_o, we randomly generated multiple starting points for the coexistence system and count how many iterations it takes for the system to reach a stable point. The ECDF of the counted number of iterations is illustrated in Figure 5.2. We can see that in a practical case where N_o is not large, the number of iterations per offerer is small enough. When online auction is enabled in Algorithm 3, the average frequency of online winning (i.e., case a in section 5.4) that occurs in a convergence process is shown in Table 5.1. We can see that the durations of considerable iterations can be shorten, since waiting time for an online iteration is always shorter than that for a regular iteration, which is equal to the duration of an entire auction period.

Second, we evaluate the spectrum utilization of proposed framework. Because spectrum reuse is achieved by a multi-winner auction at each offerer, spectrum utilization can be characterized by the number of winners per offerer, whose ECDF is illustrated in Figure 5.3

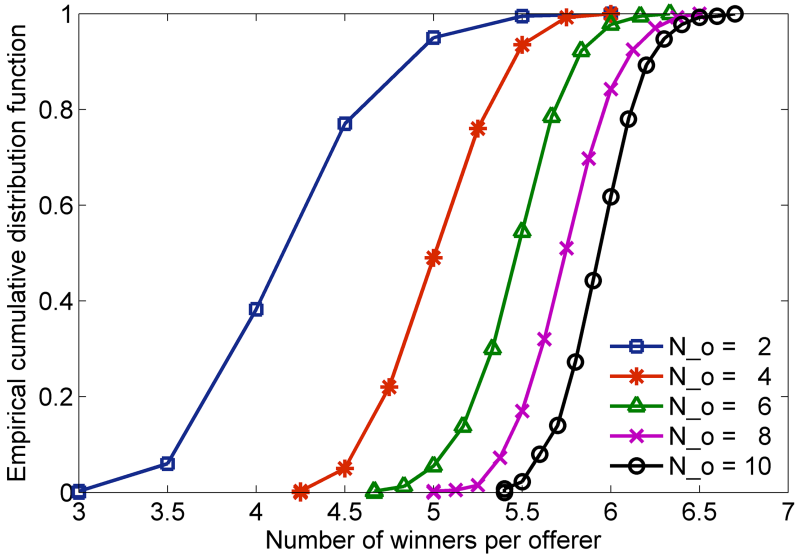

Figure 5.3: Spectrum utilization of CTSE framework.

for certain fixed N_o. We can see that the proposed framework always outperforms regular CT-CXP that only allows one winner at a time. In the proposed coexistence framework, however, each offerer can accommodate 5 to 6 requesters on average at the same time so as to achieve more efficient spectrum sharing.

Third, we evaluate the optimality gap of proposed framework. We start from discussing the total utility of requesters, i.e., U_o^L, and the total revenue, i.e., Q_o, achieved at each offerer o. The average values of U_o^L and Q_o (excluding extra bonus $\tilde{p}_o$) in a convergence process are illustrated in Figure 5.4. We can see that U_o^L is always increasing before reaching an upper bound, which verifies Lemma 2 and Lemma 3. Because the offerer's pricing strategy is defined by (5.3) and (5.10) here, Q_o drops to zero in the end due to the leave of losing requesters. However, each offerer can still collect considerable credit tokens by accommodating multiple winners, compared to the average spectrum valuation '0.5'. The optional bonus of credit tokens can further improve offerer's revenue by a fraction of maximized social welfare. Other definitions of pricing functions can also be developed to improve offerer's revenue. Although it has been proved in Theorem 1 that the coexistence system always converges to a stable equilibrium solution in a distributed manner, such a distributed solution may not be unique

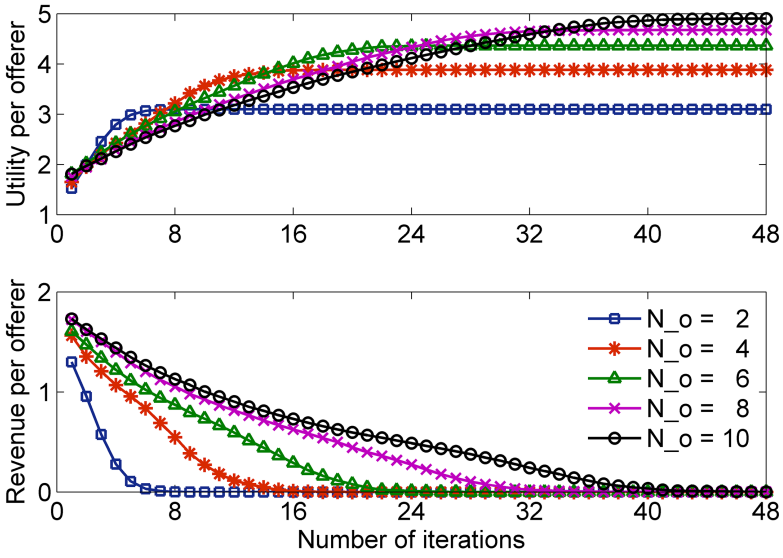

Figure 5.4: Utility and revenue per offerer in a convergence process.

and can be a local optimal solution due to multiple possible stable points. Hence, we need to verify the optimality of distributed solution compared to the centralized solution, which is given by a certain centralized infrastructure that maximizes total social welfare, i.e., $\sum_{o\in\mathcal{O}} S_o$, globally. The ratio of our distributed solution to its impractical centralized counterpart in terms of $\sum_{o\in\mathcal{O}} S_o$ is illustrated in Figure 5.5 for certain fixed N_o. We can see that the stable points given by our distributed system are usually very close to the optimal point.

Fourth, we evaluate the negative impact of control message collisions. Here, the common control channel is accessed by IEEE 802.11 CSMA/CA MAC. The average increase in the number of iterations and the average decrease in social welfare caused by control message collisions in a convergence process are shown in Table 5.2. We can see that the number of iterations increases largely, but the convergence of coexistence framework is not harmed due to retransmission and rebidding. The average value of social welfare slightly decreases owing to the collisions of expected winners' bidding messages.

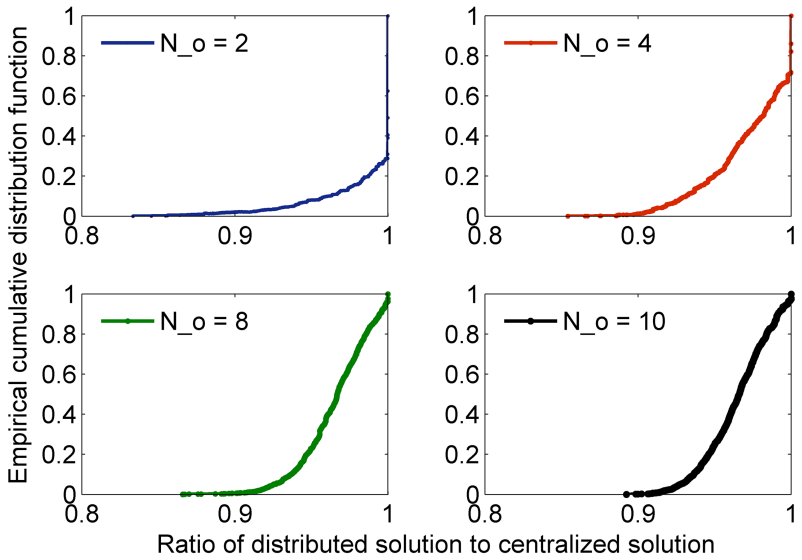

Figure 5.5: Optimality gap of CTSE framework.

5.9 Chapter Summary

In this chapter, we have addressed spectrum sharing among heterogeneous CR networks with equal priority through the CTSE framework. Specifically, the game-auction coexistence framework has been proposed in order to extend IEEE 802.16h CT-CXP in many aspects, including optimization trade-offs, spectrum reuse, online auction winning, bidding truthfulness, and system stability. Our simulation results have shown that the proposed coexistence framework always converges to a near-optimal distributed solution without a centralized coexistence infrastructure, and it improves heterogeneous coexistence jointly in truthfulness, coexistence fairness, and spectrum utilization.

Table 5.1: Frequency of online winning.

num. offerers	4	6	8	10
frequency	24.85%	30.58%	36.02%	38.89%

Table 5.2: Impact of control message collisions.

num. offerers	4	6	8	10
num. iterations	51.82%	48.53%	43.45%	42.59%
social welfare	-4.35%	-3.07%	-2.59%	-2.02%

Chapter 6

Mobility Support in Database-Driven Opportunistic Spectrum Access

In this chapter, we focus on the fourth scenario to study database-driven OSA, when secondary users are mobile [20].

6.1 Challenges and Contributions

In OSA, primary users and secondary users are usually co-located and share the same swaths of spectrum. Incumbent protection then becomes the major concern, and can be addressed by spectrum sensing techniques and/or geolocation database services. Protecting incumbent users while effectively identifying white spaces is challenging when spectrum sharing relies solely on sensing-only CRs. For this reason, database-driven OSA [4, 5, 6] has found favor with the spectrum regulators and the wireless industry. Henceforth, we assume a database-driven spectrum sharing system that is consistent with the SAS for spectrum sharing in the 3.5 GHz band [18, 19], which has recently been proposed by the FCC. A geolocation database has access to the operating characteristics of incumbent users, such as service types, channel

reservations, and protection requirements. In database-driven OSA, a registered secondary user (or its home base station) exchanges information with the geolocation database through a database query-and-update procedure. On the one hand, the user queries the database by its location information and retrieves local white space availability. On the other hand, the user updates the database with its operating information, including its location, so that the updated knowledgebase can facilitate not only incumbent protection but also location-aided network coexistence [7, 18, 19, 8].

The above database-driven spectrum sharing system can work well with static secondary networks. However, when the secondary users are *mobile*, the problem of incumbent protection becomes very challenging [5, 86]. Furthermore, the problem of network coexistence, i.e., spatial spectrum reuse and mutual interference mitigation among mobile secondary users, is even more difficult, and has not been addressed in the existing work. When supporting mobile users, each user can query and update the database at a certain interval, and the *database query interval* (or location update interval) is the key operational parameter that impacts the performance of the spectrum sharing system [19]. On the one hand, during these intervals, the database may have to rely on past location information to ensure incumbent protection and mitigate mutual interference. This leads to considerable loss of white space opportunities [5] and increased likelihood of suboptimal resource allocations. On the other hand, it can be unacceptable to blindly reduce the intervals because of limited database processing capacity, significant communication overhead rise, and needless network resource waste. This causes the database to become a bottleneck in processing spectrum sharing information. The goal of this chapter is to study the strategies for determining and adapting the database query intervals to strike an appropriate trade-off between interference-aware spectrum sharing and cost-effective database access.

We focus on two major problems for supporting secondary user mobility in database-driven OSA: white space allocation (WSA) and location update control (LUC). The database performs WSA and centrally allocates white spaces for secondary users based on their probabilistic *location uncertainty* levels aiming to mitigate co-channel interference among them

and ensure no interference with primary users. Each mobile user performs LUC and locally adjusts its location uncertainty level by adapting its database query interval, so that an appropriate trade-off between interference mitigation effectiveness and database query cost can be achieved. The two problems are solved alternately via the control of location uncertainty.

Formulating a solution approach that jointly solves these two problems is challenging. First, the location uncertainty level of a user has to be quantified to model WSA and LUC. Second, the LUC design for each user has to locally adapt database query interval to the time-varying spectrum environment without global knowledge. Third, the WSA design for the database has to promptly adjust resource allocation in response to the dynamic changes in spectrum availability, location uncertainty, and mutual interference probability.

To address the above challenges, the contributions of this chapter are summarized as follows.

- We propose a probabilistic coexistence framework, which incorporates WSA at the database and LUC at the users. This framework jointly addresses the issues of spectrum sharing and database access in database-driven OSA.
- For WSA, we derive a centralized real-time solution that minimizes the probability of mutual interference among secondary users and guarantees full protection of primary users. For LUC, we design a local two-level strategy that minimizes the weighted sum of mutual interference probability and database query frequency. This strategy combines movement-driven and interference-driven control of location uncertainty.
- We evaluate the algorithms for WSA and LUC in two types of simulation experiments. In the first experiment, we have simulated WiFi-like secondary networks using a theoretical mobility model. In the second experiment, we have simulated cellular-like secondary networks using a real-world mobility trace dataset. Our results suggest that our framework determines and adapts the database query intervals of mobile users to realize near-optimal interference mitigation with minimal location updates.
- We provide a summary of discussions on the system design guidelines and real-world deployment issues in database-driven OSA.

The remainder of this chapter is organized as follows. Related work is discussed in section 6.2. The probabilistic coexistence framework is described in section 6.3. The problems of WSA and LUC are formulated and solved in sections 6.4 and 6.5, respectively. Simulation results are presented in section 6.6. Implementation discussions are given in section 6.7. The chapter is summarized in section 6.8.

6.2 Related Work

Recently, the FCC issued a NPRM for enabling small cell use in the 3.5 GHz band [18, 19]. In the proposed framework, spectrum sharing can be achieved among three tiers of users, including tier-1 incumbent access users, tier-2 priority access users, and tier-3 general authorized access users. The tier-1 and tier-2 users are protected in certain exclusion zones, while tier-3 users have to opportunistically access the band. All the users register with the SAS that incorporates a geo-location database and various interference mitigation techniques. The SAS needs to i) specify appropriate operations across the tiers of users based on location-specific data on spectrum occupancy; ii) resolve interference issues promptly; and iii) coordinate the registered users, if necessary, based on user-generated data on spectrum sharing. The SAS is able to collect location information of registered secondary users and achieve location-aided spectrum sharing.

The research on supporting secondary user mobility in OSA is still in its infancy. The problem of protecting primary users from being interfered with by mobile secondary users has been studied in [5, 86]. In [5], the authors propose a service called SenseLess, which is a database-driven white space network. To enable mobile users without loss of white space opportunities, the database query interval of a user traveling at 60 miles/hour is suggested to be shorter than 30 seconds. However, such a high database query frequency may not always be necessary. When the database centrally manages resource allocation, a certain loss of white spaces is acceptable as long as the remaining white spaces can be enough

to offer what the user has requested. For a practical database query strategy, one should consider the trade-off between interference mitigation effectiveness and database query cost. Furthermore, the authors oversimplify the way of setting a database query interval—i.e., they propose that the database query interval of a user is inversely proportional to the speed of the user. For a cost-effective database query strategy, however, one should consider the behaviors of all the primary users and coexisting secondary users in the vicinity to determine the interval in a smart way. In [86], the authors propose enabling mobile users in a sensing-only CR network. To protect primary users under the location uncertainty of a secondary user, a guard distance is controlled to enlarge each primary exclusion zone for an extra protection. In database-driven spectrum sharing, however, it would be more challenging to achieve the coexistence of mobile secondary users due to the need for dynamic control of both interfering and interfered users' location uncertainty levels.

In CR networks without the issue of location uncertainty, the problems of joint resource allocation have been studied in [13, 17, 70, 87, 88]. In our problem, however, we try to find out how certain a location estimate should be to jointly optimize spectrum sharing and database access, which is more challenging.

6.3 System Model

In this section, we introduce our probabilistic coexistence framework and explain the basic assumptions and basic ideas.

6.3.1 Database-Driven Spectrum Sharing

We assume a spectrum sharing system similar to that in the 3.5 GHz band [18, 19]. A set of primary users $\mathcal{P}$ and a set of secondary users $\mathcal{S}$ are co-located and share the same set of channels $\mathcal{K}$. The primary users are protected from any harmful interference in certain

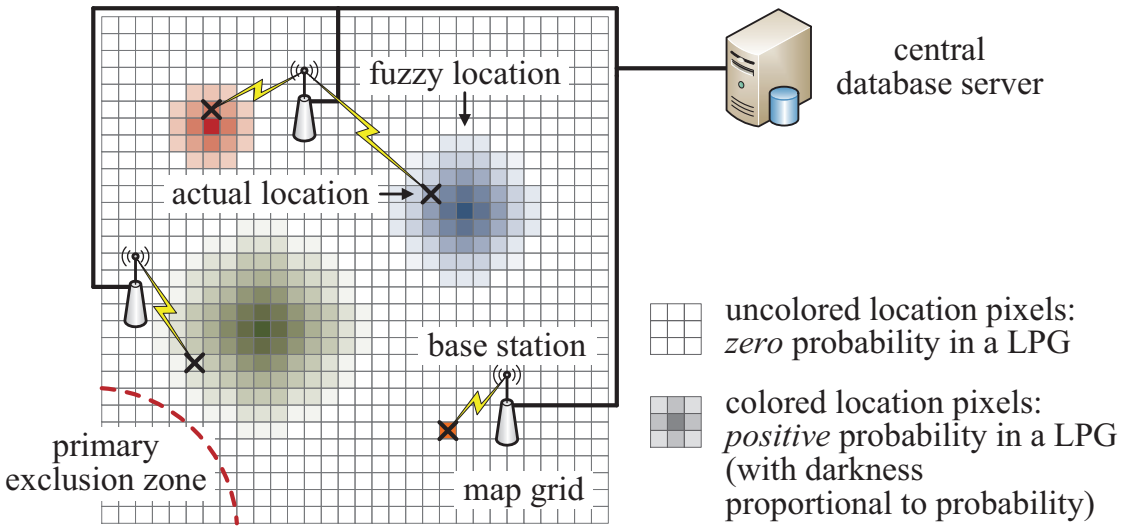

Figure 6.1: An example of database-driven spectrum sharing.

pre-defined exclusion zones, while the secondary users have to tolerate harmful interference from others. All the users register with a central database server similar to the SAS in the 3.5 GHz band, which operates a geolocation database with the database query-and-update functionality and performs WSA for the registered users. In this work, the terms "central database server" and "database" will be used interchangeably. The base stations offer the interfaces between the registered users and the database. Each user $s \in \mathcal{S}$ can be mobile, and performs LUC for itself. At a certain interval, it updates the database with its current operating information, including its location, and queries the database for location-specific resource allocation. We assume that the users do not share location information among each other for privacy protection and overhead reduction. We assume a time-slot-based system. At the beginning of each time slot, location updates are submitted from the users according to the decisions for LUC in the previous time slot. Using the location updates, the database makes a decision for WSA and notifies the users of changed resource allocation when necessary through their home base stations. Note that even if a user does not query the database, the database still computes resource allocation based on its estimation of the user's location. In this work, a database query is equivalent to a location update. A user accesses the database at most once in each time slot.

6.3.2 Location Probability Grid

The issues of incumbent protection and spectrum reuse in database-driven spectrum sharing rely on the location information of secondary users. When the users are mobile, their locations can become uncertain and thus affect resource allocation decisions. Hence, a database-driven system needs to quantify the location uncertainty level of a user between successive location update events so that the decisions for WSA and LUC can be adjusted accordingly. In this work, we leverage location probability grids (LPGs) to achieve this. In the current designs of geolocation databases, spectrum availability is computed for each location pixel in a pre-defined map grid $\mathcal{L}$ (e.g., a map grid with the granularity of 50m×50m pixel). Hence, as illustrated in Figure 6.1, we can define a LPG for each mobile user as the probability distribution of the user's location estimate over the map grid $\mathcal{L}$. For each user $s \in \mathcal{S}$, its LPG is defined by $\mathcal{Q}_s(t) \triangleq \{q_{s,l}(t) \text{ for } l \in \mathcal{L}\}$, where each $q_{s,l}(t)$ is the probability that user s is at location pixel l in time slot t. The LPG of each user defines the fuzziness of its location estimate at the database, which is a function of the elapsed time since the latest location update. The *fuzzy location* (or called movement contour) of user s is defined by $\mathcal{L}_s(t) \triangleq \{l \,|\, q_{s,l}(t) > 0 \text{ for } l \in \mathcal{L}\}$, which includes all the possible locations of the user in time slot t. Let $\ell_s(t)$ denote the *actual location* of user s in time slot t, which may not be known by the database. The database has to perform WSA based on the knowledge of LPGs. Each user needs to perform LUC to update its LPG when necessary and make its fuzzy location capture its actual location. The problems of WSA and LUC interact via the control of LPGs.

6.3.3 A Probabilistic Coexistence Framework

Intuitively, there exists a trade-off between interference mitigation effectiveness and database query cost. On the one hand, frequent database access decreases the fuzziness of location estimates and keeps the system deterministic, so that the database can precisely allocate white spaces without losing spectrum opportunities or creating mutual interference. But frequent database access may not always be necessary. For example, when a user occupies

a channel exclusively or the user is far away from other co-channel users, fuzzy location of the user is acceptable due to low likelihood of mutual interference. Moreover, too frequent database access may cause delay in server response time or introduce other types of cost such as battery power drain. On the other hand, infrequent database access increases the fuzziness of location estimates, so that large user movement contours may lead to overconservative resource allocation and thus low spectrum utilization. Hence, an optimal database query frequency needs to be found to address the trade-off.

The above trade-off issue involves two problems: (i) WSA, i.e., spectrum sharing controlled by the database; and (ii) LUC, i.e., database access controlled by the mobile users. For WSA, the database first generates LPGs based on the current and past location updates, and then minimizes the probability of mutual interference among secondary users while guaranteeing full protection of primary users based on the updated LPGs. For LUC, the users determine whether to submit location updates next to minimize the weighted sum of mutual interference probability and database query frequency based on the local measurements of user movement and co-channel interference. The problems are solved alternately through the control of LPGs. In the next two sections, the two problems are addressed separately.

6.4 White Space Allocation

In this section, we formulate the problem of WSA and then provide a centralized solution.

6.4.1 Problem Formulation

In each time slot t, the central database server updates its estimation of all the registered users' LPGs based on their current and past location updates. Using the updated LPGs, the database performs WSA to minimize the probability of mutual interference among secondary users under the constraint that the users do not cause harmful interference with primary

users. The database achieves this by coordinating the channel assignments and transmit power levels of secondary users. The decision for WSA can be broadcasted to the users through their home base stations.

For WSA that mitigates mutual interference, we first create a model of likely interference among secondary users with fuzzy locations. For each user $s \in \mathcal{S}$ and each channel $k \in \mathcal{K}$, define $X_{s,k}(t)$ as a binary channel allocation indicator such that $X_{s,k}(t) = 1$ (0) represents user s is (is not) allocated to channel k in time slot t. Due to spatial spectrum reuse, denote the set of coexisting users that share the same channel k in time slot t by $\mathcal{S}_k(t) \triangleq \{s \,|X_{s,k}(t) = 1 \text{ for } s \in \mathcal{S}\} = \{s_1, ..., s_{M_k}\}$. A possible location distribution of the co-channel users is $(l_1, ..., l_{M_k}) \in \mathcal{L}^{M_k}$. Given $\mathcal{Q}_{s_m}(t)$ for all $s_m \in \mathcal{S}_k(t)$, the probability that a particular location distribution $(l_1, ..., l_{M_k})$ happens is $\prod_{m=1}^{M_k} q_{s_m,l_m}(t)$. Thus, the database computes the probability of interference experienced at each user s_m on channel k in time slot t by

$$I_{s_m,k}(t) \triangleq \sum_{(l_1,...,l_{M_k}) \in \mathcal{L}^{M_k}} J_{s_m|l_1,...,l_{M_k}}(t) \prod_{m=1}^{M_k} q_{s_m,l_m}(t), \tag{6.1}$$

where $J_{s_m|l_1,...,l_{M_k}}(t)$ is a binary interference indicator such that $J_{s_m|l_1,...,l_{M_k}}(t) = 1$ (0) represents certain intolerable interference is (is not) experienced at user s_m for the location distribution $(l_1, ..., l_{M_k})$. Given a certain $(l_1, ..., l_{M_k})$, the database can utilize a radio wave propagation model to compute the SINR at the receiver of link s_m, denoted by $R_{s_m|l_1,...,l_{M_k}}(t)$. Here, link s_m refers to the link in which user s_m can be either a transmitter (in the uplink) or a receiver (in the downlink). We consider basic path loss model as an example, and other models can also be applied in our framework. Let

$$R_{s_m|l_1,...,l_{M_k}}(t) = \frac{Y_{s_m,k}(t) H_{l_m,l_m}(t)}{\sum_{\substack{m'=1 \\ m' \neq m}}^{M_k} Y_{s_{m'},k}(t) H_{l_{m'},l_m}(t) + Z_0}, \tag{6.2}$$

where $Y_{s_m,k}(t)$ denotes the transmit power from link s_m on channel k, $H_{l_m,l_m}(t)$ denotes the transmission gain in link s_m, $H_{l_{m'},l_m}(t)$ denotes the interference gain from link $s_{m'}$ to link

s_m, and Z_0 denotes noise power. Assume a threshold $\check{R}_{s_m}$ such that harmful interference to link s_m occurs if $R_{s_m|l_1,\ldots,l_{M_k}}(t) < \check{R}_{s_m}$. Let

$$J_{s_m|l_1,\ldots,l_{M_k}}(t) = \mathbf{1}\left(R_{s_m|l_1,\ldots,l_{M_k}}(t) < \check{R}_{s_m}\right), \tag{6.3}$$

where $\mathbf{1}()$ is an indicator function that is equal to 1 (0) if its condition is true (false). Combining (6.1), (6.2), and (6.3), $I_{s,k}(t)$ is a function of $\mathcal{Q}_s(t)$, $X_{s,k}(t)$, $Y_{s,k}(t)$ for $s \in \mathcal{S}$.

Having formulated $I_{s,k}(t)$ under location uncertainty, we can now formulate the problem of WSA. Given $\mathcal{Q}_s(t)$ for all $s \in \mathcal{S}$, the database needs to find $X_{s,k}(t)$ and $Y_{s,k}(t)$ for $s \in \mathcal{S}$, $k \in \mathcal{K}$ to minimize the global objective function

$$I(t) \triangleq \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} I_s(t), \tag{6.4}$$

where $I_s(t)$ is the average value of interference probability $I_{s,k}(t)$ over $K_s(t) \triangleq \sum_{k\in\mathcal{K}} X_{s,k}(t)$ channels that are taken by user s at a time, namely

$$I_s(t) = \frac{1}{K_s(t)}\sum_{k\in\mathcal{K}} X_{s,k}(t) I_{s,k}(t). \tag{6.5}$$

There are three constraints for WSA. The first constraint ensures the protection of primary users. In database-driven white space networks, primary users are protected in pre-defined exclusion zones. Hence, all the possible locations of secondary users under location fuzziness must be outside the exclusion zones of co-channel primary users. Suppose each channel k is occupied by a set of primary users $\mathcal{P}_k(t)$ and a set of secondary users $\mathcal{S}_k(t)$. The union of the exclusion zones of primary users in $\mathcal{P}_k(t)$ is defined by $\tilde{X}_{k,l}(t)$, which is a binary channel occupancy indicator such that $\tilde{X}_{k,l}(t) = 1$ (0) represents channel k is not (is) available for secondary users at location pixel l in time slot t. The fuzzy location of each secondary user s is characterized by $\mathcal{L}_s(t)$. Hence, the constraint for incumbent protection is defined as

$$q_{s,l}(t)[1 - (X_{s,k}(t) + \tilde{X}_{k,l}(t))] \geq 0 \text{ for } s \in \mathcal{S}, k \in \mathcal{K}, l \in \mathcal{L}. \tag{6.6}$$

The second constraint avoids secondary user starvation by ensuring a minimum number of channels, $\check{K}_s$, to be taken by each user s, which is able to be expressed as

$$K_s(t) \geq \check{K}_s \text{ for } s \in \mathcal{S}. \tag{6.7}$$

The third constraint sets a limit on the maximum transmit power, $\hat{Y}_s$, of each user s, which is able to be expressed as

$$Y_{s,k}(t) \leq X_{s,k}(t)\hat{Y}_s \text{ for } s \in \mathcal{S}, k \in \mathcal{K}. \tag{6.8}$$

In summary, the problem of WSA solved at the database in time slot t is written as follows.

Problem 1 (White Space Allocation)

Given: $\mathcal{Q}_s(t)$ for $s \in \mathcal{S}$;

Find: $X_{s,k}(t), Y_{s,k}(t)$ for $s \in \mathcal{S}, k \in \mathcal{K}$;

Minimize: $I(t)$;

Subject to: (6.6), (6.7), (6.8).

Because each $\mathcal{Q}_s(t)$ for computing $I(t)$ follows a general probability distribution, Problem 1 is a mixed-integer nonlinear program (MINLP), which is NP-hard in general. Furthermore, since the database needs to solve Problem 1 in every time slot, it cannot afford the time to find the real optimal solution. Therefore, we are going to extend the genetic algorithm in [70] to derive a fast heuristic solution to Problem 1.

6.4.2 A Genetic Algorithm

Due to the complexity of Problem 1 and the requirement of fast database response, we apply a variant of genetic algorithm [70] to derive a solution for WSA iteratively in real time. In Algorithm 5, line 1 updates LPGs for computing the objective value, and line 3 and lines 4-8

Algorithm 5 for WSA at the database in time slot t

compute $\mathcal{Q}_s(t)$ for $s \in \mathcal{S}$ according to past location update history submitted from user s
for each generation **do**
  clone a number of best feasible solutions from previous generation as offsprings
  select enough pairs of good feasible solutions from previous generation as parents
  for each pair of parents **do**
    reproduce a pair of offsprings via crossover, which randomly swaps some rows in matrices $[X_{s,k}(t)]_{|\mathcal{S}|\times|\mathcal{K}|}$ for parent 1 and $[X'_{s,k}(t)]_{|\mathcal{S}|\times|\mathcal{K}|}$ for parent 2 (same crossover is applied for $[Y_{s,k}(t)]_{|\mathcal{S}|\times|\mathcal{K}|}$ and $[Y'_{s,k}(t)]_{|\mathcal{S}|\times|\mathcal{K}|}$)
    mutate each offspring according to a probability, where a row in matrix $[X_{s,k}(t)]_{|\mathcal{S}|\times|\mathcal{K}|}$ is randomly reordered without loss of feasibility (same mutation is applied for $[Y_{s,k}(t)]_{|\mathcal{S}|\times|\mathcal{K}|}$)
  end for
  break when runtime reaches a threshold
end for

implement the two stages of cloning and breeding in a genetic algorithm, respectively. In each step of a generation, the candidate solutions are guaranteed to be feasible to Problem 1. A heuristic solution with the best objective value can be obtained when Algorithm 5 terminates. In section 6, we will evaluate the optimality gap of the solution given by Algorithm 5.

6.5 Location Update Control

In this section, we formulate the problem of LUC and then design a local strategy.

6.5.1 Problem Formulation

In time slot t, each mobile user $s \in \mathcal{S}$ performs LUC to adjust $\mathcal{Q}_s(t+1)$ for the next WSA by determining whether it wants to submit a location update in the next time slot $t+1$. For user s, define $A_s(t)$ as a binary database query indicator such that $A_s(t) = 1$ (0) represents user s submits (does not submit) a location update to the database in time slot t. In general, $\mathcal{Q}_s(t)$ can be generated by a movement prediction model based on the sequence of $A_s(t')$ for $t' = t, t-1, ...$ and the past location update history in the database.

For LUC that addresses cost-effective database access, we first define the weighted average database query cost (i.e., database query frequency) of mobile users in time slot t by

$$C(t) \triangleq \frac{\theta}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} A_s(t), \tag{6.9}$$

where θ is a positive weight factor, and can be set to reflect various overhead associated with processing query requests at both servers and devices. It may reflect cost factors such as computing power [5], energy consumption, control communication overhead, server response delay, and resource reallocation frequency. Expression (6.9) makes the cost function scale proportional to the frequency of database queries.

Having modeled $C(t)$, we can now formulate the problem of LUC, assuming global knowledge of $\mathcal{Q}_s(t)$, $X_{s,k}(t)$, $Y_{s,k}(t)$ for $s \in \mathcal{S}$, $k \in \mathcal{K}$ and future knowledge of $X_{s,k}(t+1)$, $Y_{s,k}(t+1)$ for $s \in \mathcal{S}$, $k \in \mathcal{K}$ at this time. In time slot t, given the solutions to Problem 1 obtained from time slot t and predicted from time slot $t+1$, the mobile users need to find $A_s(t+1)$ for $s \in \mathcal{S}$ to minimize the objective function

$$G(t+1) \triangleq \delta I(t+1) + (1-\delta) C(t+1), \tag{6.10}$$

where $I(t+1)$ is the objective function of Problem 1, and $\delta \in [0,1]$ is a weight factor that can be set on demand.

In addition to the constraints (6.6), (6.7), and (6.8), there is one more constraint for LUC, which ensures that the estimated $\mathcal{L}_s(t)$ can always capture the actual $\ell_s(t)$. We have

$$\ell_s(t) \in \mathcal{L}_s(t) \text{ for } s \in \mathcal{S}. \tag{6.11}$$

This constraint (6.11) is needed to guarantee the effectiveness of incumbent protection and interference mitigation in Problem 1. If any $\ell_s(t) \notin \mathcal{L}_s(t)$ occurs, even though constraint (6.6) makes sure that $\mathcal{L}_s(t)$ does not overlap with any primary exclusion zone, it is likely that

$\ell_s(t)$ moves into an exclusion zone, resulting in harmful interference from user s to primary users. Without constraint (6.11), it is also possible that the value of $I(t)$ computed using $\mathcal{L}_s(t)$ cannot accurately reflect the actual interference from or to user s located at $\ell_s(t)$.

In summary, the problem of LUC solved at mobile users in time slot t is written as follows.

Problem 2 (Location Update Control)

Given: $A_s(t')$ for $t' = t, t-1, ..., s \in \mathcal{S}$;
$X_{s,k}(t), Y_{s,k}(t)$ for $s \in \mathcal{S}, k \in \mathcal{K}$;
$X_{s,k}(t+1), Y_{s,k}(t+1)$ for $s \in \mathcal{S}, k \in \mathcal{K}$;

Find: $A_s(t+1)$ for $s \in \mathcal{S}$;

Minimize: $G(t+1)$;

Subject to: (6.6), (6.7), (6.8), (6.11).

If the database is responsible for solving Problem 2 centrally, constraint (6.11) cannot be guaranteed. Because user s does not have to continuously query and update the database in every time slot, the database cannot always keep track of $\ell_s(t)$. It is possible that a rapid change in a user's speed or direction makes the estimated movement contour become inaccurate and lose track of the user. This constraint, however, can be easily satisfied at a user if Problem 2 is solved locally. Each user s knows $\ell_s(t)$ (e.g., via GPS) and its past location update history. The user can locally compute $\mathcal{Q}_s(t)$ (and $\mathcal{L}_s(t)$) using the same movement prediction model as used at the database [89] and compare it with $\ell_s(t)$ to see whether constraint (6.11) holds. In the rest of this section, we will derive a local heuristic solution to Problem 2 through a local strategy for setting each $A_s(t+1)$.

6.5.2 A Local Two-Level Strategy

Solving Problem 2 locally at mobile users is challenging. Each user s cannot have the future knowledge of $\mathcal{Q}_{s'}(t+1)$ for $s' \neq s$ and $X_{s,k}(t+1)$, $Y_{s,k}(t+1)$ for $s \in \mathcal{S}$. As a heuristic, we can assume a short-term prediction based on the correlation between time slot t and time

slot $t+1$ [90] and the fact that the local decision of a user has limited impact on the global decision for all the users. We have

$$X_{s,k}(t+1) = X_{s,k}(t), Y_{s,k}(t+1) = Y_{s,k}(t) \text{ for } s \in \mathcal{S}, k \in \mathcal{K}. \tag{6.12}$$

Then, each user s only needs to set its own $A_s(t+1)$ as a part of a solution to Problem 2. However, user s cannot have the global knowledge of $\mathcal{Q}_{s'}(t)$, $X_{s',k}(t)$, $Y_{s',k}(t)$ for $s' \neq s$. Keeping track of such information locally at each user demands real-time message exchanges among all the users at all time. This is prohibitively expensive and very impractical. Thus, we need to propose a local strategy for each user s to find the proper setting of $A_s(t+1)$ that solves Problem 2 only based on local measurements.

The basic idea for our heuristic local strategy is based on locally evaluating the impact of a location update on the minimization of mutual interference $I(t+1)$ in (6.10). Local measurements of user movement and co-channel interference can be used to set $A_s(t+1)$ through a two-level strategy. On the first level, a *movement*-driven strategy (MDS) identifies the "must-update" and "no-update" instances with regard to the constraints of Problem 2. A location update is necessary if skipping it will violate constraint (6.11) or harm spectrum utilization greatly due to white space loss caused by constraint (6.6). A location update is avoidable if skipping it will not negatively affect interference mitigation and constraint satisfaction. For the instances when MDS is not sufficient to tell the impact of a location update, on the second level, an *interference*-driven strategy (IDS) further identifies the "no-update" instances from these uncertain instances with regard to the objective function of Problem 2. A location update is unnecessary in this step if doing it will not improve interference mitigation. In the rest of this section, we will first illustrate the designs of the two strategies, and then combine them as a local heuristic algorithm for LUC.

Movement-Driven Strategy

The MDS for LUC locally identifies the "must-update" and "no-update" instances with regard to the constraints of Problem 2. These instances are identified according to two design needs. First, constraint (6.11) should always hold. Second, constraint (6.6) should not lead to inefficient spectrum utilization. To explain our design formally, we will show how we mathematically quantify the location accuracy and location fuzziness of $\mathcal{Q}_s(t)$ for each user $s \in \mathcal{S}$ and how to use these two values in meeting the two needs.

We define the *location accuracy* of $\mathcal{Q}_s(t)$ by $q_s(t) \triangleq q_{s,\ell_s}(t)$, which is the probability of locating $\ell_s(t)$ by $\mathcal{Q}_s(t)$. The first need is addressed by the following two propositions.

Proposition 1 For each user $s \in \mathcal{S}$, constraint (6.11), i.e., $\ell_s(t) \in \mathcal{L}_s(t)$, can be locally guaranteed if $q_s(t) > 0$ holds.

Proof Please refer to the definition of $\mathcal{L}_s(t)$. □

In robot navigation research area, a LPG is often created by using Bayes filters, which probabilistically estimate the location of an object from noisy observations. In our problem, however, each mobile user exactly reports its location information. Hence, we only need to implement one prediction step of general Bayes filters without sequential localization error correction. In this work, we integrate grid-based localization [91] and sampling-based localization [92, 93] to generate LPGs.

Specifically, the database can utilize the sequence of $A_s(t')$ for $t' = t, t-1, ...$ and the past location update history to generate $\mathcal{Q}_s(t)$. Suppose that the latest location update from user s happens in time slot t_0, i.e., $A_s(t_0) = 1$ and $A_s(t') = 0$ for $t' = t, t-1, ..., t_0+1$. Define $\tilde{t}_s(t) = t - t_0$ as the elapsed time since the last location update. The past location update history can include the reported location $\tilde{l}_s$ and velocity vector $\tilde{\mathbf{v}}_s$ in time slot t_0. In each prediction step for locating user s using a technique of sampling-based Monte Carlo localization, all the samples for location estimation are placed at $\tilde{l}_s$ initially. Then, each

sample is shifted according to its randomly generated succeeding location

$$\dot{L}_s(t) \triangleq \tilde{l}_s + (\tilde{\mathbf{v}}_s + \tilde{\mathbf{w}}_s)\,\tilde{t}_s(t), \tag{6.13}$$

where $\dot{L}_s(t)$ is a random variable generated by the movement prediction model representing an estimate of $\ell_s(t)$, and $\tilde{\mathbf{w}}_s$ is a vector of random variables characterizing the uncertainty in generating $\mathcal{Q}_s(t)$. The $\mathcal{Q}_s(t)$ is essentially the probability distribution of $\dot{L}_s(t)$ over $\mathcal{L}$. Then using (6.13), we are able to have

$$q_s(t) = Pr\{\tilde{l}_s + (\tilde{\mathbf{v}}_s + \tilde{\mathbf{w}}_s)\,\tilde{t}_s(t) = \ell_s(t)\}. \tag{6.14}$$

The probability distribution of $\tilde{\mathbf{w}}_s$ varies with the applied movement prediction model and the knowledge of underlying terrain data. As an example, we assume $\tilde{\mathbf{w}}_s$ follows a bivariate normal distribution with parameters $(\mu_x = 0, \mu_y = 0, \sigma_x^2, \sigma_y^2, \rho)$ in the two-dimensional map grid with no terrain knowledge. Then, we have the following common results.

Proposition 2 For $\tilde{\mathbf{w}}_s$ that follows a bivariate normal distribution, $q_s(t)$ is upper bounded by a monotonically decreasing function of $\tilde{t}_s(t)$, i.e.,

$$\tilde{q}_s(t) = Pr\{\tilde{l}_s + (\tilde{\mathbf{v}}_s + \tilde{\mathbf{w}}_s)\,\tilde{t}_s(t) = E[\dot{L}_s(t)]\},$$

where $E[\dot{L}_s(t)] = \tilde{l}_s + \tilde{\mathbf{v}}_s\tilde{t}_s(t)$. As $\ell_s(t)$ deviates from $E[\dot{L}_s(t)]$, $q_s(t)$ is a monotonically decreasing function of $\tilde{t}_s(t)$.

Proof Let $\tilde{\mathbf{w}}_s = (W_x, W_y)$ and define a vector of random variables $(L_x, L_y) = \tilde{t}_s(W_x, W_y) + (\tilde{l}_x, \tilde{l}_y)$, where $(\tilde{l}_x, \tilde{l}_y) = \tilde{l}_s + \tilde{\mathbf{v}}_s\tilde{t}_s$ can be viewed as a constant vector. Based on the probability density function (PDF) of (W_x, W_y) and the properties for linear transformations of bivariate random variables, the joint PDF of (L_x, L_y) can be derived as

$$f_{L_x,L_y}(l_x, l_y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\frac{1}{\tilde{t}_s^2} \times \exp\left(-\frac{\left(\frac{l_x-\tilde{l}_x}{\sigma_x}\right)^2 + \left(\frac{l_y-\tilde{l}_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{l_x-\tilde{l}_x}{\sigma_x}\right)\left(\frac{l_y-\tilde{l}_y}{\sigma_y}\right)}{2(1-\rho^2)}\frac{1}{\tilde{t}_s^2}\right). \tag{6.15}$$

Hence, $q_s(t)$ is the double integral of $f_{L_x,L_y}(l_x,l_y)$ over the small area of location pixel $\ell_s(t)$, and is a function of $\tilde{t}_s(t)$ for $\tilde{t}_s(t)>0$. The size of $\ell_s(t)$ is fixed, so $q_s(t)$ is only determined by $f_{L_x,L_y}(l_x,l_y)$. For $\rho\in(-1,1)$, we know that $\frac{\left(\frac{l_x-\tilde{l}_x}{\sigma_x}\right)^2+\left(\frac{l_y-\tilde{l}_y}{\sigma_y}\right)^2-2\rho\left(\frac{l_x-\tilde{l}_x}{\sigma_x}\right)\left(\frac{l_y-\tilde{l}_y}{\sigma_y}\right)}{2(1-\rho^2)}\geq 0$. Hence,

$$f_{L_x,L_y}(l_x,l_y)\leq\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\frac{1}{\tilde{t}_s^2}. \tag{6.16}$$

The equality in (6.16) holds when $(l_x,l_y)=(\tilde{l}_x,\tilde{l}_y)$. In this perfect case, $\ell_s(t)$ can always be captured by $E[\dot{L}_s(t)]$, which is on the top of the bell-shaped joint density $f_{L_x,L_y}(l_x,l_y)$. In this case, $q_s(t)$ reaches its upper bound $\tilde{q}_s(t)$, which is the double integral of $f_{L_x,L_y}(l_x,l_y)$ over $\ell_s(t)=E[\dot{L}_s(t)]$. Since $\tilde{q}_s(t)$ is determined by $f_{L_x,L_y}(l_x,l_y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\frac{1}{\tilde{t}_s^2}$, $\tilde{q}_s(t)$ is a monotonically decreasing function of $\tilde{t}_s(t)$.

As $\ell_s(t)$ deviates from $E[\dot{L}_s(t)]$, $d_{\ell_s,\tilde{l}_s}(t)<d_{\ell_s,\tilde{l}_s}(t+1)$, where $d_{\ell_s,\tilde{l}_s}(t)$ is the distance between $\ell_s(t)$ and $E[\dot{L}_s(t)]$. It is easy to show $q_s(t)>q_s(t+1)$ through (6.15). □

Even though $q_s(t)$ may have local fluctuations when $\tilde{t}_s(t)$ increases due to likely irregular trajectory of user s, it is bounded by the monotonically decreasing upper bound and it is common that $\ell_s(t)$ deviates from $E[\dot{L}_s(t)]$ without localization correction. Its general trend is decreasing with time. Hence, setting $A_s(t+1)=0$ will mostly lead to a further decrease in $q_s(t)$. To ensure $q_s(t+1)>0$, user s needs to track the drop in $q_s(t)$ and set $A_s(t+1)=1$ when $q_s(t)$ is too low. This addresses the first need for constraint (6.11).

We define the *location fuzziness* of $\mathcal{Q}_s(t)$ by $|\mathcal{L}_s(t)|$, which is the covered pixel area of $\mathcal{L}_s(t)$. The second need can be addressed by the following proposition.

Proposition 3 For $\tilde{\mathbf{w}}_s$ that follows a bivariate normal distribution, $|\mathcal{L}_s(t)|$ is a monotonically increasing function of $\tilde{t}_s(t)$.

Proof The size of $|\mathcal{L}_s(t)|$ relies on the covariance matrix of $(L_x,L_y)=\tilde{t}_s(W_x,W_y)+(\tilde{l}_x,\tilde{l}_y)$,

$$\begin{bmatrix}(\tilde{t}_s\sigma_x)^2 & \rho(\tilde{t}_s\sigma_x)(\tilde{t}_s\sigma_y)\\ \rho(\tilde{t}_s\sigma_x)(\tilde{t}_s\sigma_y) & (\tilde{t}_s\sigma_y)^2\end{bmatrix},$$

where both $\tilde{t}_s\sigma_x$ and $\tilde{t}_s\sigma_y$ grow with $\tilde{t}_s(t)$. Hence, for a fixed ρ, it is straightforward to show that $|\mathcal{L}_s(t)|$ enlarges as $\tilde{t}_s(t)$ increases. □

Hence, setting $A_s(t+1) = 0$ will lead to a further increase in $|\mathcal{L}_s(t)|$, which indicates that $\mathcal{L}_s(t+1)$ is a fuzzier estimate of $\ell_s(t+1)$ compared with $\mathcal{L}_s(t)$ as an estimate of $\ell_s(t)$. To avoid significant loss of white space opportunities and high chance of inaccurate resource allocations, user s needs to set $A_s(t+1) = 1$ when necessary. This addresses the second need for not being affected by constraint (6.6) too much.

If $\tilde{\mathbf{w}}_s$ follows a general probability distribution, different movement prediction models can achieve different levels of location estimation deviation, but the properties of lower location accuracy and greater location fuzziness are usually true as the elapsed time increases.

Local MDS The MDS works as follows. Each user s can track $q_s(t)$ and set $A_s(t+1)$ using two thresholds: α_1 and α_2 ($\alpha_1 \geq \alpha_2$). There are three possible cases:

a. When $q_s(t) \geq \alpha_1$, user s assumes "no-update" case and sets $A_s(t+1) = 0$.

b. When $\alpha_2 < q_s(t) < \alpha_1$, user s assumes uncertain case and needs to consider using IDS for further investigation.

c. When $q_s(t) \leq \alpha_2$, user s assumes "must-update" case and sets $A_s(t+1) = 1$.

For case a, a relatively small $|\mathcal{L}_s(t)|$ ensures that $\mathcal{L}_s(t)$ accurately captures $\ell_s(t)$ and thus constraints (6.11) and (6.6) are easy to satisfy. If user s sets $A_s(t+1) = 1$, it is very likely that this action has little impact on the resource allocation decision under the slightly reduced location uncertainty. For case c, a relatively large $|\mathcal{L}_s(t)|$ tends to cause significant loss of white space opportunities due to constraint (6.6) and high chance of inaccurate interference estimation. If user s sets $A_s(t+1) = 0$, it is very likely that the database makes a suboptimal resource allocation decision under the increased location uncertainty. More importantly, if $q_s(t)$ is very low, $q_s(t+1) = 0$ is possible and thus violates constraint (6.11). For case b, IDS is needed as an additional step to set $A_s(t+1)$.

Interference-Driven Strategy

The IDS for LUC further locally identifies the "no-update" instances among the uncertain instances from MDS (in case b) with regard to the objective function of Problem 2. To decide whether $A_s(t+1)$ should be 0 or 1, user s needs to predict which action will lead to a smaller objective value. Ideally, user s can take the following process. First, user s solves Problem 1 for time slot $t+1$ to obtain the values of $I(t+1)$ for both $A_s(t+1)=0$ and $A_s(t+1)=1$ cases. Next, user s compares the values of $G(t+1)$ for the two cases. Whichever objective value that comes the smallest corresponds to the optimal setting of $A_s(t+1)$ for Problem 2. However, this process requires user s to have perfect global knowledge, which is impractical to obtain. Thus, we need to derive a heuristic strategy to set $A_s(t+1)$, which can approximate this process but only requires local knowledge.

Based on (6.9) and (6.10), we are able to know that the key in the design of IDS is to find a local way to determine whether

$$I(t+1)|_{A_s(t+1)=0} - I(t+1)|_{A_s(t+1)=1} > \frac{(1-\delta)\theta}{\delta|\mathcal{S}|}. \tag{6.17}$$

If (6.17) is likely to be true (false), set $A_s(t+1)=1$ (0).

Global IDS To find a way to judge (6.17), we consider that user s using a channel $k \in \mathcal{K}$ has a set of co-channel neighbors $s' \in \mathcal{S}_k(t)$, $s' \neq s$. As an example, we discuss IDS with global knowledge based on the co-channel interference in the uplink from user s to others s'. Define $d_{s'}(t)$ as the protection radius from the receiver of each interfered link s', and define $d_{s,s'}(t)$ as the distance between the transmitter of the interfering link s and the receiver of each interfered link s'. Define $\omega_1 \in [0,1)$ and $\omega_2 \in (1,\infty)$. With proper ω_1 and ω_2, there are three possible relative location cases:

1. When $d_{s,s'}(t) \leq \omega_1 d_{s'}(t)$ for any $s' \in \mathcal{S}_k(t)$, $s' \neq s$, user s assumes "no-update" case and sets $A_s(t+1)=0$.
2. When $\omega_1 d_{s'}(t) < d_{s,s'}(t) < \omega_2 d_{s'}(t)$ for any $s' \in \mathcal{S}_k(t)$, $s' \neq s$, user s sets $A_s(t+1)=1$.

3. When $d_{s,s'}(t) \geq \omega_2 d_{s'}(t)$ for all $s' \in \mathcal{S}_k(t)$, $s' \neq s$, user s assumes "no-update" case and sets $A_s(t+1) = 0$.

For case 1, user s is deep inside the protection area of a link s'. Because MDS guarantees that $\ell_s(t) \in \mathcal{L}_s(t)$ and $|\mathcal{L}_s(t)|$ is not too large in case b, there should be a relatively full overlap between $\mathcal{L}_s(t)$ and the protection area of link s'. Thus, even setting $A_s(t+1) = 0$, Problem 1 is likely to judge that links s and s' have a high probability of mutual interference and make a resource allocation decision as same as that in the case of setting $A_s(t+1) = 1$ (mutual interference probability is 1). Because the decisions for WSA in both update and no-update cases should be the same, (6.17) is likely to be false. For case 3, user s is far outside the protection area of any link $s' \in \mathcal{S}_k(t)$. The movement contour of user s is likely to be mostly outside the protection area of any link s'. Thus, even setting $A_s(t+1) = 0$, Problem 1 is likely to judge that links s and s' have a low probability of mutual interference and make a resource allocation decision as same as that in the case of setting $A_s(t+1) = 1$ (mutual interference probability is 0). Because the decisions for WSA in both update and no-update cases should be the same, (6.17) is likely to be false. For case 2, user s is near the boundary of the protection area of a link s'. The movement contour of user s is partly inside and partly outside the protection area of link s'. If user s sets $A_s(t+1) = 1$, Problem 1 can precisely know whether there is mutual interference between links s and s'. If there exists mutual interference, the database will try to reallocate channels to links s and s' to eliminate the interference. If there exists no mutual interference, the database is likely to keep the interference to be zero. In both cases, the solution to Problem 1 usually makes the probability of mutual interference between links s and s' be zero. If user s sets $A_s(t+1) = 0$, Problem 1 can only use the movement contour to obtain a positive value of mutual interference probability. The left-hand side of (6.17) is usually large, so (6.17) is likely to be true.

Unfortunately, user s is not able to precisely know which case it belongs to without perfect knowledge of other co-channel users' location information, which is impractical to obtain. However, user s can estimate which relative location case it is likely in based on its local measurement of $R_{s,k}(t)$, the SINR that user s has observed on channel k in time slot t.

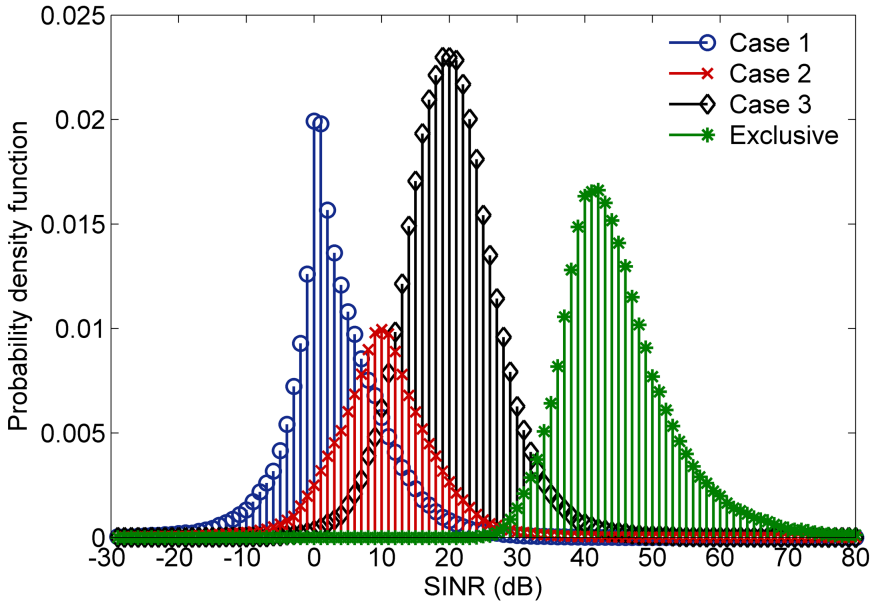

Figure 6.2: PDFs of $R_{s,k}(t)$ ($\omega_1 = \frac{2}{3}$, $\omega_2 = \frac{4}{3}$).

Proposition 4 The relative location cases and the local measurement of $R_{s,k}(t)$ are correlated in a stochastic sense.

For example, we use numerical results to create Figure 6.2, which shows the PDFs of $R_{s,k}(t)$ observed in different relative location cases. The exclusive case refers to the situation where there is no any other user sharing the channel with user s. Obviously, a location update is not needed in the exclusive case. The PDFs of the other three cases are averaged when the number of co-channel users ranges from 2 to 7. The details of the 7-cell system will be presented in section 6. We observe that when the locally measured $R_{s,k}(t)$ is either low enough or high enough, the false judgement rates for the "no-update" cases (except for case 2) are relatively low due to their distinct peaks in the PDFs. Then, user s can usually identify the "no-update" instances properly.

Local IDS The IDS works as follows. Each user s that takes a channel k can measure $R_{s,k}(t)$ and set $A_s(t+1)$ using two thresholds: β_1 and β_2 ($\beta_1 \leq \beta_2$). There are three possible cases mapping with the relative location cases:

Algorithm 6 for LUC at user $s \in \mathcal{S}$ in time slot t

if $A_s(t) == 1$ (0) **then**
  query (do not query) the database
end if
access white space spectrum according to $X_{s,k}(t)$, $Y_{s,k}(t)$ for $k \in \mathcal{K}$ assigned by the database
compute $\mathcal{Q}_s(t)$ as in Algorithm 5, and measure $q_s(t)$
if case a (c) **then**
  set $A_s(t+1) = 0$ (1)
else
  set $A_s(t+1) = 0$
  measure $R_{s,k}(t)$ for all k with $X_{s,k}(t) == 1$
  if case 2' for any k **then**
    set $A_s(t+1) = 1$
  end if
end if

1'. When $R_{s,k}(t) \leq \beta_1$, user s assumes "no-update" case and sets $A_s(t+1) = 0$.

2'. When $\beta_1 < R_{s,k}(t) < \beta_2$, user s sets $A_s(t+1) = 1$.

3'. When $R_{s,k}(t) \geq \beta_2$, user s assumes "no-update" case and sets $A_s(t+1) = 0$.

For case 1', user s is most likely in case 1. For case 3', user s is very likely in either case 3 or the exclusive case. For case 2', user s can be in case 1, 2, or 3. Using a conservative strategy that cares more about interference mitigation effectiveness rather than database query cost, user s has to assume the worst case 2.

Combining MDS and IDS

The local two-level strategy for LUC combining MDS and IDS is summarized in Algorithm 6. Two sets of thresholds, i.e., movement-related α_1 and α_2, and interference-related β_1 and β_2, are used to identify the cases where a location update is unnecessary (in cases a, 1', 3') or necessary (in cases c, 2'). Running Algorithm 6 at mobile users gives a heuristic solution to Problem 2. After the thresholds have been determined, the complexity of Algorithm 6 mainly comes from computing $\mathcal{Q}_s(t)$ through one prediction step of a Bayes filter, which is not very expensive. In sections 6.6 and 6.7, we will discuss how these thresholds can be set.

6.6 Performance Evaluation

In this section, we evaluate the proposed algorithms using the results collected from two simulation experiments. First, we have simulated WiFi-like secondary networks using a theoretical mobility model. Second, we have simulated cellular-like secondary networks using a real-world mobility trace dataset.

6.6.1 Model-Driven Simulation

In the first experiment, we study a 7-cell spectrum sharing system, in which the seven base stations for secondary networks are placed at the center and six vertices of a regular hexagon with the edge length of 1500 m. Total m channels are shared by m primary users and $2m$ secondary users. Each primary user occupies a different channel in its exclusion zone with the radius of 1500 m, and its duty cycle is λ. The primary users are uniformly distributed. Each secondary user is randomly associated with a base station, and let $\check{K}_s = 1$, $\hat{Y}_s = 100$ mW, $\check{R}_s = 10$ dB. The random movement of each user is defined by a semi-Markov smooth mobility model [94]. The target direction in the model is restricted back when the user is more than 1500 m away from its home base station, and the average target speed is on the order of human walking speed. The granularity of the map grid is 50m×50m, and the duration of a time slot is 120 s. In this example, the cells largely overlap with each other, and user mobility relative to the size of map grid is high enough to generate dynamic inter-cell interference. We explain the settings of α_1, α_2 in MDS and β_1, β_2 in IDS to adapt the database query intervals. The average objective function is recorded as $\bar{G} = \delta\bar{I} + (1-\delta)\bar{C}$, where $\bar{I}$ is the average fraction of time for unsuccessful reception in each link and $\bar{C}$ is the average value of $C(t)$ over time with $\theta = 1$.

First, we show the impact of α_1 and α_2 on $\bar{G}$. In Figure 6.3, fixing $\alpha_2 = 0.01$, $\beta_1 = 5$ dB, $\beta_2 = 14$ dB, we increase α_1 from α_2 to 1. This increases the probability of case 2' in IDS and thus increases $\bar{C}$. In Figure 6.4, fixing $\alpha_1 = \alpha_2$, we increase α_2 from 0 to 1. This increases

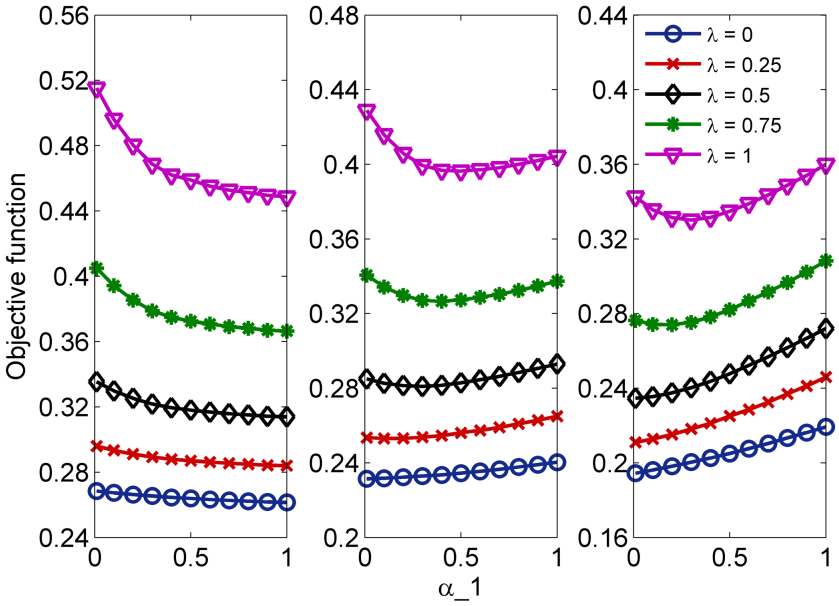

Figure 6.3: Impact of α_1 on objective $\bar{G}$ ($m = 28$): a) $\delta = 1$ (left); b) $\delta = 0.8$ (middle); c) $\delta = 0.6$ (right).

the probability of case c in MDS and thus increases $\bar{C}$. To balance between $\bar{I}$ and $\bar{C}$, larger α_1 and larger α_2 are chosen for larger δ to keep higher location accuracy and smaller location fuzziness. Larger $\alpha_1 - \alpha_2$ is used to give more preference to IDS.

Second, we show the impact of β_1 and β_2 on $\bar{G}$. From Figure 6.2, too small β_1 or too large β_2 can create a lot of unnecessary location updates due to falsely treating case 1 or 3 as case 2'. On the contrary, too large β_1 or too small β_2 can omit a lot of necessary location updates due to falsely treating case 2 as case 1' or 3'. The values of β_1 and β_2 should be chosen with regard to the false judgement rates for the relative location cases. In Figure 6.5, fixing $\alpha_1 = 0.7$, $\alpha_2 = 0.01$, we increase $\beta_2 - \beta_1$ from 0 (decreasing β_1 from 9.5 dB and increasing β_2 from 9.5 dB). This increases the probability of case 2' in IDS and thus increases $\bar{C}$. To balance between $\bar{I}$ and $\bar{C}$, smaller β_1 and larger β_2 are chosen for larger δ to reduce the false judgements that cause missing necessary location updates, while larger β_1 and smaller β_2 are chosen for smaller δ to reduce the false judgements that cause unnecessary location updates.

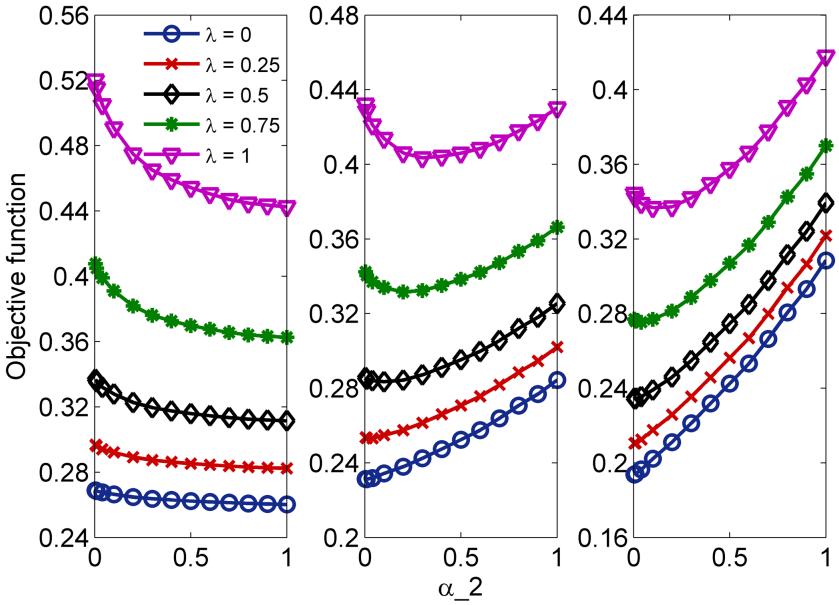

Figure 6.4: Impact of α_2 on objective $\bar{G}$ ($m = 28$): a) $\delta = 1$ (left); b) $\delta = 0.8$ (middle); c) $\delta = 0.6$ (right).

Third, we compare the value of $1 - \bar{I}$ given by our heuristic mechanism with the value given by an impractical optimal mechanism. Here, our mechanism is called movement-driven and interference-driven database query strategy (MIDQ). We also consider two alternative heuristic mechanisms: a pure movement-driven database query strategy (MDQ) (by fixing $\alpha_1 = \alpha_2$) and a periodic database query strategy (PDQ) (as used in [5]). The optimal mechanism assumes perfect global knowledge, so that all the users' actual locations are exactly known and no any location update is needed. Then, the optimal mechanism performs a centralized resource allocation accordingly based on the accurate location information (without the consideration of location update cost). The ratio of $1 - \bar{I}$ from a heuristic solution to that from the optimal solution is depicted in Figure 6.6. It can be seen that the ratio grows with increasing database query frequency. For MIDQ, $\bar{C}$ is increased by first increasing α_1 ($\alpha_2 = 0.01$, $\beta_1 = 5$ dB, $\beta_2 = 14$ dB) from 0.01 to 1 and then increasing $\beta_2 - \beta_1$. For MDQ, $\bar{C}$ is increased by increasing $\alpha_1 = \alpha_2$ from 0.01 to 1. Compared with MDQ and PDQ, the proposed MIDQ achieves better reduction of mutual interference with the same

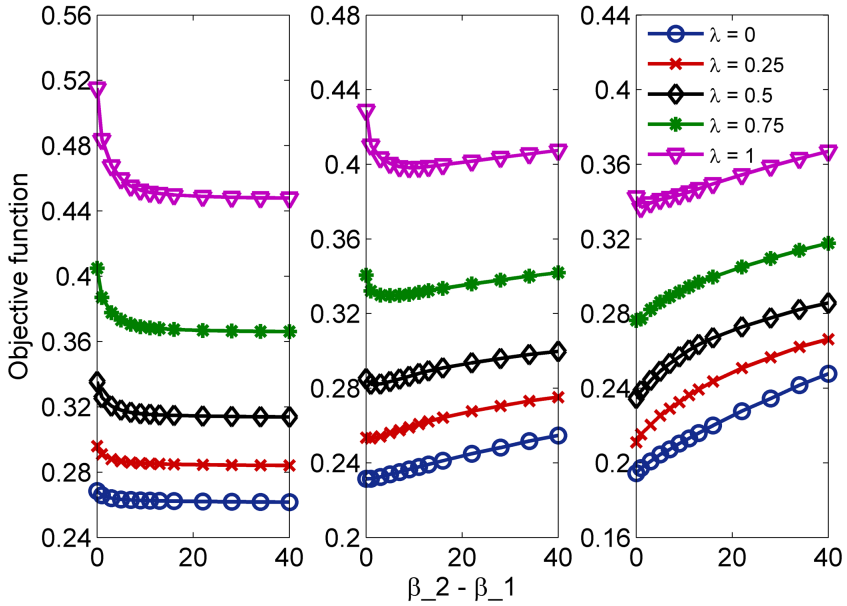

Figure 6.5: Impact of β_1 and β_2 on objective $\bar{G}$ ($m = 28$): a) $\delta = 1$ (left); b) $\delta = 0.8$ (middle); c) $\delta = 0.6$ (right).

level of database query frequency (cost). With a high enough database query frequency, MIDQ can achieve almost the same interference level as the optimal mechanism.

6.6.2 Trace-Driven Simulation

In the second experiment, we have utilized a real-world mobility trace dataset that contains the GPS coordinates of real taxis collected in the San Francisco Bay Area [95]. In this experiment, a taxi represents a mobile secondary user. However, we need to deploy the base stations for secondary networks, since the dataset does not include any location information of cellular base stations or WiFi access points. We use a similar method as used in [96]. The possible locations of the base stations to be deployed are the location points that have been traveled by the taxis in the dataset, and the points are along roads and around buildings. Each point is assigned with a selection probability, which is the frequency at which the location point has been visited by the taxis. In this way, more base stations are deployed

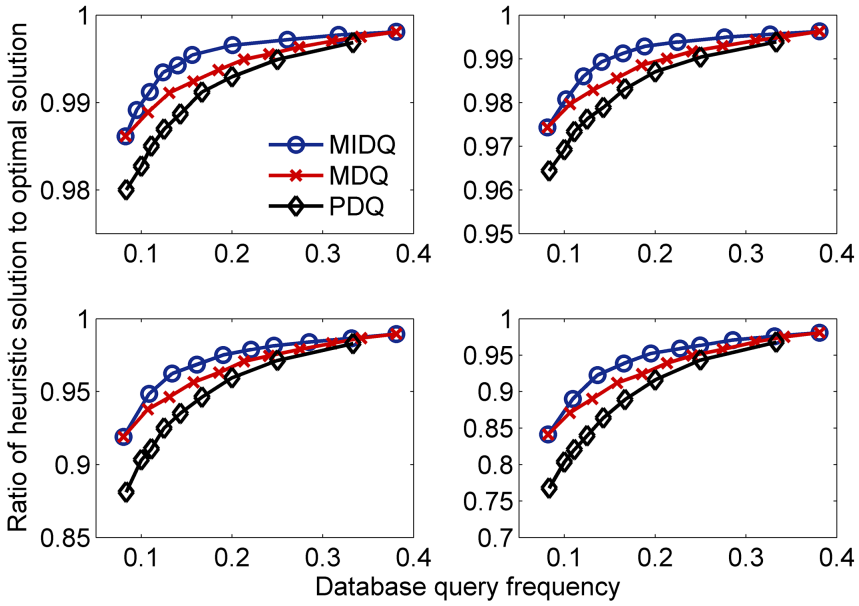

Figure 6.6: Optimality gap of coexistence framework (model-driven, $m = 28$): a) $\lambda = 0$ (top-left); b) $\lambda = 0.25$ (top-right); c) $\lambda = 0.75$ (bottom-left); d) $\lambda = 1$ (bottom-right).

in more heavily traveled areas. Total m channels are shared by m primary users and $2m$ secondary users. As shown in Figure 6.7, the traces of 48 taxis are selected, and 12 base stations are deployed. We assume that each taxi is associated with the nearest base station, i.e., inter-cell handover is enabled. The exclusion zones of the primary users with a radius of 3000 m are uniformly distributed in the simulation area. The granularity of the map grid is 100m×100m, and the duration of a time slot is 60 s. Similar to Figure 6.6, the ratio of the heuristic solution to the optimal solution in terms of $1 - \bar{I}$ is shown in Figure 6.8. From the results, we can see that the proposed MIDQ still outperforms MDQ and PDQ. The overall performance of the heuristic mechanisms degrades in this experiment compared with that in the first experiment. Besides the impact of incumbents, this should be caused by the high-speed vehicular mobility that leads to less accurate movement prediction and the dynamic user-cell association that leads to less accurate local interference estimation.

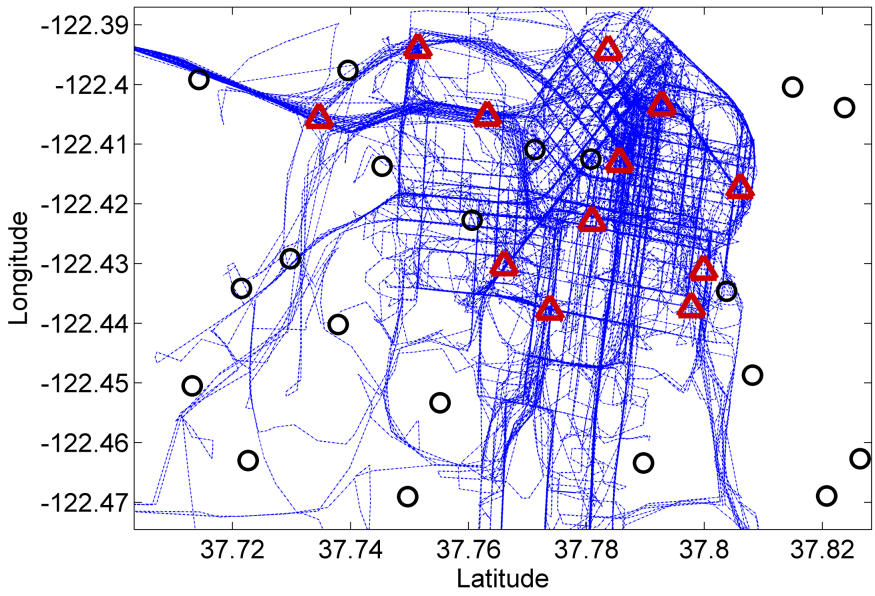

Figure 6.7: A trace-driven simulation scenario in San Francisco (lines: taxi movement trajectories; triangles: secondary base stations; circles: primary exclusion zone centres).

6.7 Guidelines and Discussions

In this section, we generally discuss the system design guidelines and real-world deployment issues in database-driven OSA.

1 The database query interval/frequency of each mobile user should be adapted according to both internal factors (e.g., user mobility, database processing capacity, and location privacy) and external factors (e.g., incumbent protection, and network coexistence) in order to achieve a proper balance between spectrum sharing and database access. In general, a higher frequency of database queries is necessary if i) user movement pattern is more dynamic; ii) interference mitigation effectiveness is of higher priority than database query cost; iii) location information is less sensitive at the database; iv) the duty cycle of primary users is larger, i.e., less channels are available for certain users; v) mutual interference is more likely to occur, i.e., more users share certain channels.

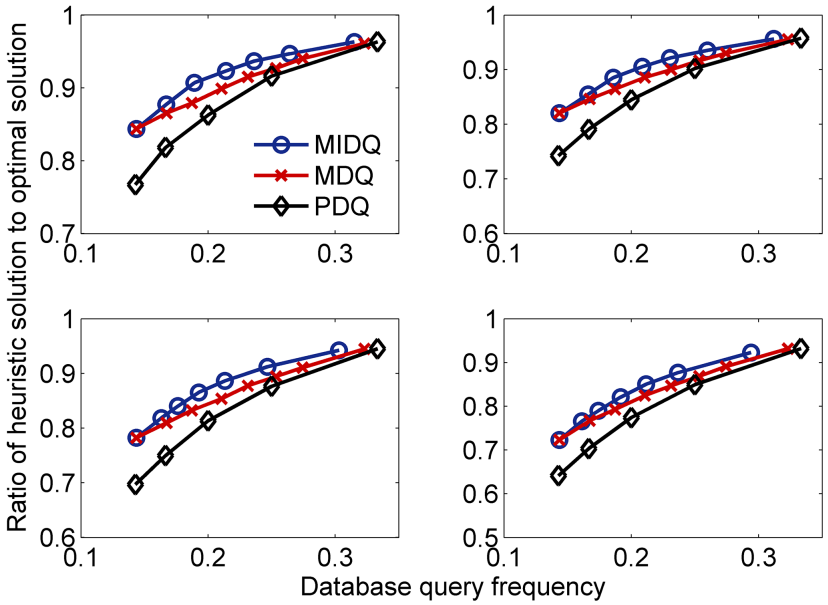

Figure 6.8: Optimality gap of coexistence framework (trace-driven): a) $\lambda = 0$ (top-left); b) $\lambda = 0.25$ (top-right); c) $\lambda = 0.75$ (bottom-left); d) $\lambda = 1$ (bottom-right).

2 In a probabilistic coexistence framework as proposed in this work, the database can set the system parameters offline (e.g., via statistical analysis) or online (e.g., via a stochastic decision process). For example, the thresholds α_1 and α_2 can be empirically set according to the maximum and minimum desired levels of location accuracy or location fuzziness, respectively, for a movement prediction model. The thresholds β_1 and β_2 can be dynamically adjusted according to the changes in the spectrum environment (e.g., primary user activity, secondary user density and mobility, and resource reallocation) and the results from past decisions (e.g., the rate at which erroneous decisions were made in determining appropriate case for a user). To reduce complexity, the thresholds can also be empirically set according to the PDFs as in Figure 6.2. One can learn the amount of false judgement for each case given a particular setting of β_1 and β_2, and thus make an educated choice. The offline and online decisions can be combined to mediate a trade-off between performance and complexity.

3 For the estimation of user movement pattern at the database, the applied localization method or movement prediction model determines how the location uncertainty level of a

user changes with time. In general, a more advanced localization technique (e.g., a method that integrates GPS, maps, and user behaviors) is able to achieve a slower drop in location accuracy and a slower rise in location fuzziness, and requires less location updates for localization error correction.

4 For the estimation of harmful interference to primary users and mutual interference among secondary users at the database, the applied radio wave propagation model affects the effectiveness of spectrum availability prediction and resource allocation. In general, a more sophisticated signal propagation model (e.g., Longley-Rice with terrain data [5]) achieves better incumbent protection and higher spectrum utilization, and even further improves local decisions at the users for mutual interference mitigation.

5 It is hard to quantify the cost of database queries in database-driven spectrum sharing. However, one of major bottlenecks in such a system is the capacity of the central database server. In a fine-grained system, each single user performs LUC, and the central server manages spectrum sharing information and performs WSA at the user-level granularity. The computational complexity of resource allocation is affected by the number of registered users and the spatial granularity of map grid, and the resource allocation has to meet the timely requirement if a time slot is very short. Hence, when designing a database-driven spectrum sharing system, the maximum number of registered users that can be served by each server needs to be considered. Alternatively, each base station can manage the users in its cell as a whole. In such a coarse-grained system, each base station performs LUC, and the central server offers the services at the cell-level granularity. In this approach, each base station further performs WSA at the user-level granularity. In general, the fine-grained system is recommended when database processing capacity is not a problem, while the coarse-grained system is preferred when server response delay or resource reallocation frequency is too high.

6.8 Chapter Summary

In this chapter, we have addressed two major problems for supporting mobile users in database-driven OSA: WSA and LUC. In the proposed coexistence framework, WSA is performed centrally, while LUC is performed locally. We have used a LPG to quantify the location uncertainty level of each user. The problems of WSA and LUC have been jointly solved through dynamic control of LPGs. The local strategy for adapting database query intervals combines practical MDS and IDS to reduce database query cost without negatively impacting interference mitigation effectiveness. Our model-driven and trace-driven simulation results have shown that the intervals can be adapted to yield near-optimal interference mitigation with minimal location updates.

Chapter 7

Conclusions

In this dissertation, we have proposed a taxonomy of heterogeneous coexistence mechanisms, and then have addressed four typical coexistence scenarios for shared spectrum access: (i) sensing-based spectrum sharing when channel aggregation is employed; (ii) co-channel self-coexistence when coordination is not required; (iii) co-channel heterogeneous coexistence when coordination is allowed; and (iv) database-driven spectrum sharing when secondary user mobility is considered. We have provided models and solutions for the four scenarios.

In chapter 3, for the first scenario, we have proposed a channel usage model to analyze the impact of both primary and secondary user behaviors on the efficiency of channel aggregation. This model is general to capture a wide range of user behaviors. We have modeled the delay costs for performing channel aggregation. Our numerical and simulation results have shown that user demands in both the frequency and time domains should be carefully chosen to minimize expected cumulative delay.

In chapter 4, for the second scenario, we have proposed an uplink soft frequency reuse technique to enable globally power-efficient and locally fair spectrum sharing. We frame the self-coexistence problem as a non-cooperative game, and have designed a local heuristic algorithm that achieves the Nash equilibrium in a distributed manner. Our simulation results

have shown that the proposed technique is mostly near-optimal and improves self-coexistence in spectrum utilization, power consumption, and intra-cell fairness.

In chapter 5, for the third scenario, we have proposed a credit-token-based spectrum etiquette framework, which is a game-auction coexistence framework that enables spectrum sharing among spectrum offerers and spectrum requesters. We have proved that the framework is stable. Our simulation results have shown that the proposed framework always converges to a near-optimal distributed solution and improves coexistence fairness and spectrum utilization.

In chapter 6, for the fourth scenario, we have proposed a probabilistic coexistence framework that supports mobile users by locally adapting their location uncertainty levels in order to find an appropriate trade-off between interference mitigation effectiveness and location update cost. Our simulation results have shown that the proposed framework can determine and adapt the database query intervals of mobile users to achieve near-optimal interference mitigation with minimal location updates.

The previous body of our research mainly focuses on secondary-secondary coexistence for general shared spectrum access scenarios. In our future work, we are going to study both primary-secondary and secondary-secondary spectrum sharing in a more realistic scenario—i.e., the database-driven spectrum sharing system in the 3.5 GHz band [18, 19]. One of major problems with the development of such a system is that the incumbent exclusion zones defined by the FCC are too large so that most densely populated areas in the United States are not available for spectrum sharing. Hence, we are interested in the method for dynamically shrinking the overconservative incumbent exclusion zones to enhance spectrum utilization efficiency in these areas. One promising idea is to integrate spectrum sensing into database-driven spectrum sharing, so that we can take advantage of both the geo-location database that offers reliable incumbent characteristics and the spectrum sensing that captures dynamic spectrum environment. This idea will be discussed in the next chapter.

Chapter 8

Future Work — Integration of Spectrum Sensing into Database-Driven Spectrum Sharing

In this chapter, we explain the idea of a hybrid spectrum sharing architecture, which allows the geolocation database to incentivize spectrum sensing to dynamically shrink the over-conservative incumbent exclusion zones and enhance spectrum utilization efficiency.

8.1 Dynamic Zoning

In database-driven spectrum sharing, incumbent protection should be ensured in incumbent exclusion zones, which are defined by the geolocation database using radio wave propagation models and terrain data. Because the operating characteristics of incumbent users, such as locations, spectrum reservations, and protection requirements, are usually quasi-static, the definitions of such exclusion zones are not subject to frequent changes. In order to fully protect incumbent users, the database has to assume the worst-case interference from

secondary users and conservatively defines the exclusion zones. However, overconservative exclusion zones can be very large and thus greatly reduce the white space opportunities for spectrum sharing among secondary users, especially in densely populated areas.

To discover the "true" white spaces and improve spectrum utilization efficiency when the incumbent exclusion zones are overly conservative, we aim to find the way of dynamically shrinking the exclusion zones without negatively impacting incumbent protection. A recent study has shown that spectrum sensing is helpful to identify extra white spaces in indoor environments [97], and this observation motivates the integration of spectrum sensing into database-driven spectrum sharing to achieve our objective. In such a hybrid OSA architecture, geolocation database offers the primary solution for spectrum sharing, while spectrum sensing supplements it as the secondary solution. The construction of a radio environment map (REM) [98] has involved fusing the incumbent registrations from the database and the spectrum measurements from dedicated sensors or white space devices. Using such data fusion, we can take advantage of both the geolocation database that offers reliable incumbent characteristics and the spectrum sensing that captures dynamic spectrum environment.

8.1.1 A Three-Zone Architecture

We propose a dynamic zoning framework, which addresses the problem of utilizing spectrum sensing to refine the incumbent exclusion zones originally given by propagation modeling only. In this framework, we define three types of zones: *hard exclusion zone* (HEZ), *soft exclusion zone* (SEZ), and *unlimited access zone* (UAZ), as illustrated in Figure 8.1. The HEZ is the zone where secondary access is completely prohibited, since secondary users have very high chance of causing harmful interference with incumbents. The SEZ is the zone where secondary access is possible, but secondary users have to provide spectrum sensing results to show that the spectrum is locally available for them. The zone outside the exclusion zones is UAZ where secondary access is always allowed, since secondary users have very low chance of causing harmful interference with incumbents. In this three-zone architecture, spectrum

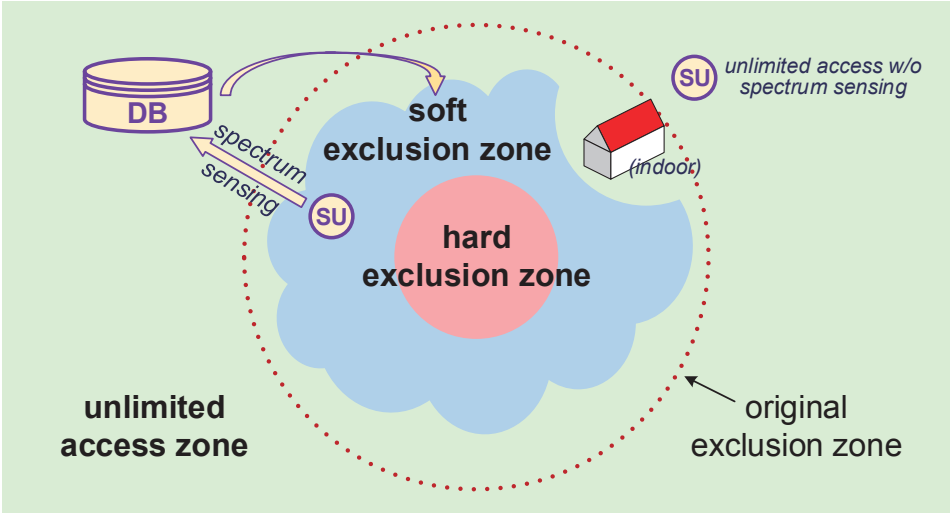

Figure 8.1: An example of dynamic zoning.

sensing is most useful in SEZ for the discovery of extra (especially indoor) white spaces that are not identified by propagation modeling, and spectrum sensing does be required for spectrum access in SEZ to guarantee incumbent protection. Compared with the quasi-static boundary of original exclusion zone, the outer boundaries of HEZ and SEZ should be dynamically refined by the spectrum sensing results collected from SEZ.

8.1.2 Zone Boundary Refining

In the above three-zone architecture, the two zone boundaries, i.e., the boundary between HEZ and SEZ, and the boundary between SEZ and UAZ, need to be defined based on the fusion of modeled incumbent characteristics and spectrum sensing results.

The outer boundary of HEZ should be less dynamic than that of SEZ, since it is harder to find extra white spaces near HEZ. Furthermore, spectrum sensing only occurs in SEZ, and no user has incentive to sense in HEZ where secondary access is always prohibited. Hence, no white space can be discovered in HEZ, and the size of HEZ is non-decreasing. In other words, depending on whether white spaces can be found right outside HEZ, spectrum sensing can either keep HEZ (if found) or enlarge HEZ (if not found). Note that the purpose of

enlarging HEZ is to enhance incumbent protection (or cut useless spectrum sensing) instead of improving spectrum utilization. To reduce complexity, the outer boundary of HEZ can only be determined by propagation modeling.

The outer boundary of SEZ should be more dynamic, since there are much more white space opportunities near UAZ. Contrary to the case of HEZ, spectrum sensing is not required for spectrum access in UAZ, and the size of SEZ is non-increasing. In other words, depending on whether white spaces can be found right inside SEZ, spectrum sensing can either keep SEZ (if not found) or shrink SEZ (if found). Hence, the outer boundary of SEZ can be refined by spectrum sensing to better support spectrum sharing.

In this chapter, we focus on the problem of *outer boundary refining* for SEZ, which is key to achieve gain in spectrum utilization efficiency. Before addressing the zoning problem, the database has to solve two other problems: *sensing incentivizing* and *data fusion*, which answer the questions of how spectrum sensing results are collected and processed, respectively. The problem of sensing incentivizing is solved to give reward to the users who have reported useful sensing information for dynamic zoning to compensate their sensing cost and incentivize high-quality spectrum sensing. The problem of data fusion is solved to integrate the two sources of incumbent information, i.e., geolocation database and spectrum sensing, to provide evidence for the new white space opportunities. The two problems will be discussed in the next two sections in detail. The operations of database and user for addressing the zoning problem are summarized in Table 8.1.

After the problems of sensing incentivizing and data fusion have been solved, the solution to the problem of outer boundary refining for SEZ can be obtained, which finds out where spectrum sensing is needed and where is not. Only the locations that have strong evidence to show they are always safe for incumbents, such as some indoor locations, can be moved from SEZ to UAZ, and spectrum sensing is no longer required for spectrum access at these locations for a period of time. For the locations that are shown to be temporarily available for spectrum sharing, it is better to keep these locations in SEZ at this time to further

Table 8.1: Operations of database and user for dynamic zoning.

	Database	User
HEZ	Outer boundary refining ← data fusion (sensing from SEZ)	Nothing
SEZ	Outer boundary refining ← sensing incentivizing, data fusion	Rewarded access if performing spectrum sensing
UAZ	Nothing	Unlimited access

incentivize spectrum sensing. The size of SEZ is non-increasing, so these locations cannot be moved back to SEZ once they become a part of UAZ. Hence, SEZ may not be the smaller the better, even though data fusion results suggest that SEZ can be smaller at this time. An appropriate size of SEZ is determined by balancing the trade-off between incumbent protection and spectrum utilization.

The dynamic zoning can be carried out iteratively. Initially, the outer boundaries of HEZ and SEZ are computed by propagation modeling. In each iteration, the database first solves the problem of sensing incentivizing to collect desired amount of spectrum sensing results. After the targeted sensing accuracy has been reached, the database solves the problem of data fusion to evaluate the probability of interfering with incumbents for the areas where sensing results are reported. Based on the current and past evaluations of aggregated interference from secondary users, the database is able to find a solution to the problem of outer boundary refining. The newly refined HEZ and SEZ will be used for zoning in the next iteration.

8.2 Sensing Incentivizing

The general idea of sensing incentivizing is similar to participatory sensing [99, 100] or mobile crowdsourcing [101, 102]. A principal-agent model is used to describe sensing incentivizing.

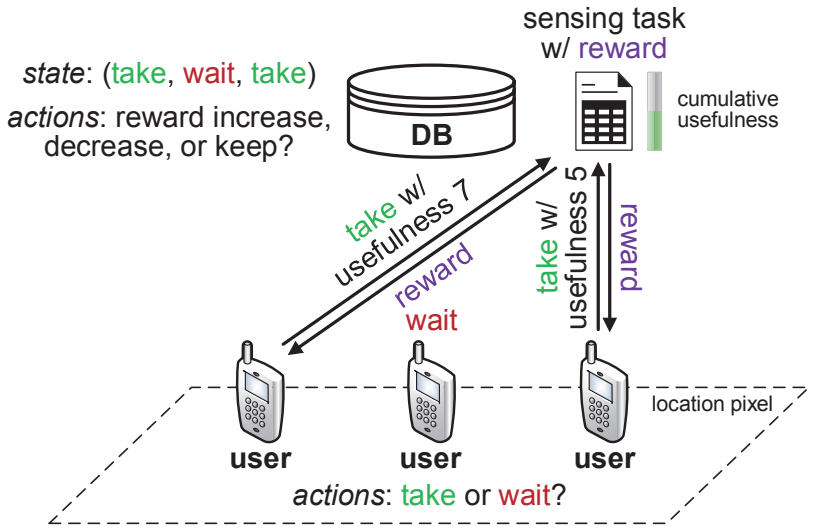

Figure 8.2: An example of sensing incentivizing.

Targeting each unit area in SEZ, the database (as the principal) announces a sensing task to support dynamic zoning. Based on the need for spectrum access, each user in the area (as an agent) autonomously responds to participate in or ignore the task. An example of sensing incentivizing is illustrated in Figure 8.2, which will be explained in this section.

8.2.1 Reward and Usefulness

Due to the involved cost for completing a sensing task, such as the time and power for sensing, a rational user has no incentive to take the task for free. Hence, some rewards should be given to incentivize the users with sensing capability to willingly and truthfully report their spectrum sensing results to the database. In existing work, the cost for sensing is usually quantified by sensing time, which is directly related to the loss of transmission time or capacity. Fortunately, each user performs spectrum sensing for spectrum access in SEZ, and the white space capacity returned to the user can be treated as its reward for sensing. Note that a correct report that the sensed spectrum is unavailable should still be rewarded by some capacity from another segment of spectrum or another period of time in the future. If a sensing report is rewarded only when the sensed spectrum is available, the users have

incentive to lie when the spectrum is actually unavailable, since the user gets nothing for an unavailable spectrum report. Hence, when the sensed spectrum is shown to be unavailable, the rewarded capacity has to come from another place in the frequency or time domain.

To guarantee the quality of spectrum sensing results for dynamic zoning, each sensing task should be assigned with a usefulness requirement. There are several factors affecting the usefulness of a sensing report. First, usefulness depends on the applied sensing technique and spent sensing time. Different types of sensing techniques along with different durations of sensing time achieve different levels of sensing accuracy (probabilities of miss detection and false alarm) and sensing sensitivity (ability of detecting weak signals). Second, usefulness depends on the location where spectrum sensing is performed relative to the area targeted by the sensing task. Even if the targeted area is not sensed, interpolation methods can be used to estimate the spectrum environment there based on the sensing results from surrounding sensing points. But the locations of those sensing points should be close enough given their sensing sensitivity levels. Third, usefulness depends on the current progress of sensing task completion. The sensing results reported when the database is in the dark about the targeted area are usually more valuable than that reported when the database has got a clear picture of the area. Fourth, usefulness depends on the evaluation of the user's past sensing reports in terms of quality and truthfulness, especially when the database has not collected enough sensing results to judge the quality and truthfulness of the current sensing report. At the initial stage of a sensing task, the database should prefer to hear from more trustworthy users. The overall usefulness of a sensing report should be a weighted sum of the evaluation scores for the above factors. The weight values may change with time.

8.2.2 Dynamic Pricing

The amount of rewarded capacity for each sensing report should be determined. Even though each sensing report comes with a usefulness score, the reward for it should not directly be proportional to the claimed usefulness, since the user may lie about the first two factors

for a better usefulness score. Like the recruitment in a real job market, the reward for a sensing report can be specified in advance in a "contract" between the database and the participating user. The sensing report should meet the basic usefulness requirement, but the claimed usefulness is not used to define the reward. Instead, the amount of reward can be given by a dynamic pricing strategy [103, 104, 105], which deals with the interactions between sensing task supply and sensing participation demand. For the over-demand case, on the one hand, the rewarded capacity for a sensing report can be reduced to accommodate more participating users given limited white space capacity. For the over-supply case, on the other hand, the rewarded capacity for a sensing report has to be raised to accelerate the progress of sensing task completion before the iteration deadline. Hence, in general, the reward for sensing is decreasing as user participation demand rises and the reward for sensing is increasing as task completion progress drops.

Database's Strategy

Now we focus on dynamic pricing in one iteration for dynamic zoning. Suppose that there are a set of channels, denoted by $\mathcal{K}$. For each channel $k \in \mathcal{K}$, there are a set of location pixels in SEZ, denoted by $\mathcal{L}_k$. For each channel $k \in \mathcal{K}$ and each location $l \in \mathcal{L}_k$, define a sensing task, $\mathcal{M}_{k,l}$, which is a set of sensing measurement reports collected before the deadline. A user can submit a sensing report anytime before the deadline to receive its reward, but the time of submission determines the amount of reward to be received.

Each iteration is divided into time slots $\{1, 2, \cdots, T\}$, where T is the deadline for sensing incentivizing. In each time slot $t \in \{1, 2, \cdots, T\}$, denote $R_{k,l}(t)$ as the change of reward for contributing to $\mathcal{M}_{k,l}$, and $R_{k,l}(t) = -\Delta$, 0, or Δ represents reward decrease by Δ, keep, or increase by Δ, respectively. Then, the current reward is $\tilde{R}_{k,l}(t) = \tilde{R}_{k,l}(t-1) + R_{k,l}(t)$. Denote $\mathcal{S}_{k,l}(t)$ as the set of users who want to participate given $R_{k,l}(t)$. In general, the size of $\mathcal{S}_{k,l}(t)$ is non-decreasing with increasing $R_{k,l}(t)$. For each user $s \in \mathcal{S}_{k,l}(t)$, the database determines $X_{k,l}^{(s)}(t)$, a binary user selection indicator such that $X_{k,l}^{(s)}(t) = 1$ (0) represents

user s is (is not) selected to contribute to $\mathcal{M}_{k,l}$. Because the reward for the selected users is in the form of white space capacity and the database has to guarantee that the (worst-case) aggregated interference with incumbents is bounded, there should be limited openings for each sensing task. The budget for the capacity to be rewarded to the users in $\mathcal{S}_{k,l}(t)$ may not be limited to the sensed channel k, since there may be some other white space channels at location l. The objective of sensing incentivizing is to maximize spectrum utilization by rewarding the selected users with the selected capacity. Hence, the value of $R_{k,l}(t)$ and the values of $X_{k,l}^{(s)}(t)$ are jointly determined to maximize expected rewarded capacity

$$\bar{R}_{k,l}(t) \triangleq \sum_{\mathcal{S}_{k,l}(t)} Pr\{\mathcal{S}_{k,l}(t)|R_{k,l}(t)\} \sum_{s\in\mathcal{S}_{k,l}(t)} X_{k,l}^{(s)}(t)|_{\mathcal{S}_{k,l}(t)}\tilde{R}_{k,l}(t), \tag{8.1}$$

in which $Pr\{\mathcal{S}_{k,l}(t)|R_{k,l}(t)\}$ is the probability that a possible combination of $\mathcal{S}_{k,l}(t)$ occurs given the setting of $R_{k,l}(t)$. After the value of $R_{k,l}(t)$ and the values of $X_{k,l}^{(s)}(t)$ are determined, each selected user s needs to submit a sensing report $m_{k,l}^{(s)}(t)$ after sensing channel k at location l and receives reward $\tilde{R}_{k,l}(t)$. Denote $U_{k,l}^{(s)}(t)$ as the usefulness of $m_{k,l}^{(s)}(t)$. To achieve a certain level of sensing accuracy, the database has a target for the cumulative usefulness score from all the sensing reports for the same sensing task. The target for cumulative usefulness is similar to the perishable goods for sale before the deadline [104]. The gap towards the target can be viewed as the unsold goods left in inventory, which defines the progress of sensing task completion.

Generally, the reward for sensing is decreasing as user participation demand rises and the reward for sensing is increasing as task completion progress drops. As an example, suppose that a sensing task requires cumulative usefulness score of 10 to complete. In the first time slot, the database announces a reward value. One user is chosen from several bidding users due to incumbent protection and the user contributes usefulness 5. In the next time slot, the database decreases the reward value because of high demand in participation and nice task completion progress (5/10). The second user is chosen and it contributes usefulness 3. In the third time slot, the database further decreases the reward value for the almost

completed task (8/10), but no one bids due to too low reward. In the fourth time slot, the database increases the reward value to incentivize sensing participation. The third user is chosen and it contributes usefulness 2. After the task has been completed with acceptable sensing accuracy, the database sets the reward to zero.

In summary, in each time slot t of an iteration (before the deadline T), the database aims to

$$\begin{aligned} \text{Find:} \quad & R_{k,l}(t); \\ & X_{k,l}^{(s)}(t) \ for \ s \in \mathcal{S}_{k,l}(t); \\ \text{Maximize:} \quad & \bar{R}_{k,l}(t); \\ \text{Subject to:} \quad & incumbent\ protection\ constraint; \\ & cumulative\ usefulness\ constraint. \end{aligned}$$

The above problem can be considered as a Markov decision process (MDP), in which the state is $\mathcal{S}_{k,l}(t)$, the action is to set $R_{k,l}(t)$, the transition probability from $\mathcal{S}_{k,l}(t-1)$ to $\mathcal{S}_{k,l}(t)$ for action $R_{k,l}(t)$ is $Pr\{\mathcal{S}_{k,l}(t)|R_{k,l}(t)\}$, and the expected reward for action $R_{k,l}(t)$ is $\sum_{s\in\mathcal{S}_{k,l}(t)} X_{k,l}^{(s)}(t)|_{\mathcal{S}_{k,l}(t)}\tilde{R}_{k,l}(t)$. There exists the uncertainty on $\mathcal{S}_{k,l}(t)$ when setting $R_{k,l}(t)$, so $Pr\{\mathcal{S}_{k,l}(t)|R_{k,l}(t)\}$ is set based on estimating the possible combinations of $\mathcal{S}_{k,l}(t)$, which depend on the users' strategies. The actual combination of $\mathcal{S}_{k,l}(t)$ can be obtained after the users have responded. Then, the values of $X_{k,l}^{(s)}(t)$ can be determined.

User's Strategy

Suppose that a user s is at a location $l \in \mathcal{L}_k$ for a channel $k \in \mathcal{K}$, and thus user s is in SEZ and has the opportunity of contributing to the sensing task $\mathcal{M}_{k,l}$. In each time slot t, user s examines the announced $R_{k,l}(t)$, and determines $A_{k,l}^{(s)}(t)$, a binary task participation indicator such that $A_{k,l}^{(s)}(t) = 1$ (0) represents user s bids (does not bid) to contribute to $\mathcal{M}_{k,l}$. Because of the involved cost for sensing, a user may not always take a task. To simplify our analysis, we can assume that each user takes a sensing task at most once in each iteration.

Then, the value of $A_{k,l}^{(s)}(t)$ is determined to maximize user s's expected rewarded capacity

$$\bar{R}_{k,l}^{(s)}(t) \triangleq A_{k,l}^{(s)}(t) \sum_{\mathcal{S}'_{k,l}(t)} Pr\{\mathcal{S}'_{k,l}(t)|R_{k,l}(t)\} Y_{k,l}^{(s)}(t)|_{\mathcal{S}'_{k,l}(t)} \tilde{R}_{k,l}(t) + \left(1 - A_{k,l}^{(s)}(t)\right) \dot{R}_{k,l}(t), \quad (8.2)$$

where $\mathcal{S}'_{k,l}(t)$ is the set of other users who want to participate in $\mathcal{M}_{k,l}$ given $R_{k,l}(t)$, $Y_{k,l}^{(s)}(t)$ is the probability that $X_{k,l}^{(s)}(t) = 1$, and $\dot{R}_{k,l}(t)$ is the estimated reward in the future. If user s is myopic, i.e., user s takes the task immediately when its reward covers its cost or its utility is positive, the second part of $\bar{R}_{k,l}^{(s)}(t)$ can be ignored. However, if user s is strategic [103, 104, 105], i.e., user s is far-sighted enough to consider the future path of reward change when making decisions on sensing participation, the user's strategy can be complex. In dynamic pricing, strategic users may achieve higher rewards in the long term but also have higher chance of losing the opportunity if the task has been completed by someone else.

In summary, in each time slot t of an iteration (before the deadline T), the user s aims to

$$\begin{aligned} \text{Given:} \quad & R_{k,l}(t); \\ \text{Find:} \quad & A_{k,l}^{(s)}(t); \\ \text{Maximize:} \quad & \bar{R}_{k,l}^{(s)}(t); \\ \text{Subject to:} \quad & \textit{positive utility constraint}. \end{aligned}$$

The local decision of user s is based on the estimation of other users' decisions $\mathcal{S}'_{k,l}(t)$. The major challenge is that each user does not have knowledge of neighboring users, which are potential competitors. Even if we assume that all the users are myopic, each user still needs to estimate the probability of being selected for spectrum sensing in the current time slot.

8.3 Data Fusion

In order to support dynamic zoning, the spectrum sensing results collected by solving the problem of sensing incentivizing in the first stage should be fully utilized. In the next stage,

the problem of data fusion is solved to integrate the two sources of incumbent characteristics, i.e., geolocation database and spectrum sensing, to provide evidence for the new white space opportunities. Different from traditional cooperative spectrum sensing, such as voting, the data fusion in our problem involves both quasi-static propagation modeling and real-time spectrum measurements. How to weigh the two sources of information and convert the information to a common white space metric is a new challenging problem.

The REM introduced in [98] is a dynamic spectrum map generated by a cognitive engine, which processes the data collected from multiple sources, including the measurements from dedicated sensors or CRs and the regulations from regulatory bodies in a country. The techniques for REM construction and interference derivation can be classified into two categories: spatial statistics based approaches and transmitter location determination based approaches. The first class of approaches, such as interpolation methods, are referred to as direct approaches. Given spectrum measurements at some points, spatial statistics can be used to estimate the missing data at some other points. The measurement point density, measurement frequency, and measurement accuracy are some key factors. The second class of approaches are referred to as indirect approaches. If the information on incumbent transmitters is available, which can be provided by geolocation database, the signal strength at each location is easier to obtain. The above two categories of techniques can be used together to take advantage of both the geolocation database that offers reliable incumbent characteristics and the spectrum sensing that captures dynamic spectrum environment.

Bibliography

[1] G. Staple and K. Werbach. The end of spectrum scarcity. *IEEE Spectrum*, 41(3), Mar. 2004.

[2] S. Haykin. Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2), Feb. 2005.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer Networks*, 50(13), Sept. 2006.

[4] D. Gurney, G. Buchwald, L. Ecklund, S. Kuffner, and J. Grosspietsch. Geo-location database techniques for incumbent protection in the tv white space. In *Proc. IEEE DySPAN'08*, Oct. 2008.

[5] R. Murty, R. Chandra, T. Moscibroda, and P. Bahl. Senseless: A database-driven white spaces network. *IEEE Transactions on Mobile Computing*, 11(2), Feb. 2012.

[6] M. Barrie, S. Delaere, G. Sukareviciene, J. Gesquiere, and I. Moerman. Geolocation database beyond tv white spaces? matching applications with database requirements. In *Proc. IEEE DySPAN'12*, Oct. 2012.

[7] C. Ghosh, S. Roy, and D. Cavalcanti. Coexistence challenges for heterogeneous cognitive wireless networks in tv white spaces. *IEEE Wireless Communications*, 18(4), Aug. 2011.

[8] B. Gao, J.-M. Park, Y. Yang, and S. Roy. A taxonomy of coexistence mechanisms for heterogeneous cognitive radio networks operating in tv white spaces. *IEEE Wireless Communications*, 19(4), Aug. 2012.

[9] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary user behavior in cellular networks and implications for dynamic spectrum access. *IEEE Communications Magazine*, 47(3), Mar. 2009.

[10] B. Gao, Y. Yang, and J.-M. Park. Channel aggregation in cognitive radio networks with practical considerations. In *Proc. IEEE ICC'11*, Jun. 2011.

[11] IEEE 802.16 WG. Ieee standard for local and metropolitan area networks, part 16: Air interface for broadband wireless access systems, amendment 2: Improved coexistence mechanisms for license-exempt operation. IEEE Std., Jul. 2010.

[12] IEEE 802.22 WG. Ieee standard for information technology, wireless regional area networks (wran), part 22: Cognitive wireless ran medium access control (mac) and physical layer (phy) specifications: Policies and procedures for operation in the tv bands. IEEE Std., Jul. 2011.

[13] B. Gao, J.-M. Park, and Y. Yang. Uplink soft frequency reuse for self-coexistence of cognitive radio networks operating in white-space spectrum. In *Proc. IEEE INFOCOM'12*, Mar. 2012.

[14] B. Gao, J.-M. Park, and Y. Yang. Uplink soft frequency reuse for self-coexistence of cognitive radio networks. *IEEE Transactions on Mobile Computing*, 13(6), Jun. 2014.

[15] T. Baykas, M. Cummings, H. Kang, M. Kasslin, J. Kwak, R. Paine, A. Reznik, R. Saeed, and S. J. Shellhammer. Developing a standard for tv white space coexistence: Technical challenges and solution approaches. *IEEE Wireless Communications*, 19(1), Feb. 2012.

[16] J. W. Mwangoka, P. Marques, and J. Rodriguez. Exploiting tv white spaces in europe: The cogeu approach. In *Proc. IEEE DySPAN'11*, May 2011.

[17] B. Gao, Y. Yang, and J.-M. Park. A credit-token-based spectrum etiquette framework for coexistence of heterogeneous cognitive radio networks. In *Proc. IEEE INFOCOM'14*, Apr. 2014.

[18] FCC. Enabling innovative small cell use in 3.5 ghz band nprm & order. GN Docket No. 12-354, FCC 12-148, Dec. 2012.

[19] FCC. Wireless telecommunications bureau and office of engineering and technology call for technical papers on the proposed spectrum access system for the 3.5 ghz band. GN Docket No. 12-354, DA 13-2213, Nov. 2013.

[20] B. Gao, J.-M. Park, and Y. Yang. Supporting mobile users in database-driven opportunistic spectrum access. In *Proc. ACM MobiHoc'14*, Aug. 2014.

[21] ECMA TC48-TG1. Standard ecma-392: Mac and phy for operation in tv white space. ECMA Std., Dec. 2009.

[22] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh. White space networking with wi-fi like connectivity. In *Proc. ACM SIGCOMM'09*, Aug. 2009.

[23] A. Flores, R. Guerra, E. Knightly, P. Ecclesine, and S. Pandey. Ieee 802.11af: A standard for tv white space spectrum sharing. *IEEE Communications Magazine*, 51(10), Oct. 2013.

[24] M. I. Rahman, A. Behravant, H. Koorapaty, J. Sachs, and K. Balachandran. License-exempt lte systems for secondary spectrum usage: Scenarios and first assessment. In *Proc. IEEE DySPAN'11*, May 2011.

[25] S. Pollin, M. Ergen, A. Dejonghe, L. van der Perre, F. Catthoor, I. Moerman, and A. Bahai. Distributed cognitive coexistence of 802.15.4 with 802.11. In *Proc. IEEE CROWNCOM'06*, Jun. 2006.

[26] J. Huang, G. Xing, G. Zhou, and R. Zhou. Beyond coexistence: Exploiting wifi white space for zigbee performance assurance. In *Proc. IEEE ICNP'10*, Oct. 2010.

[27] X. Zhang and K. G. Shin. Enabling coexistence of heterogeneous wireless systems: Case for zigbee and wifi. In *Proc. ACM MobiHoc'11*, May 2011.

[28] S. M. Mishra, S. ten Brink, R. Mahadevappa, and R. W. Brodersen. Cognitive technology for ultra-wideband/wimax coexistence. In *Proc. IEEE DySPAN'07*, Apr. 2007.

[29] X. Jing and D. Raychaudhuri. Spectrum coexistence of ieee 802.11b and 802.16a networks using the cscc etiquette protocol. In *Proc. IEEE DySPAN'05*, Nov. 2005.

[30] G. P. Villardi, Y. D. Alemseged, C. Sun, C.-S. Sum, T. H. Nguyen, T. Baykas, and H. Harada. Enabling coexistence of multiple cognitive networks in tv white space. *IEEE Wireless Communications*, 18(4), Aug. 2011.

[31] R. Ferrus, O. Sallent, and R. Agusti. Interworking in heterogeneous wireless networks: Comprehensive framework and future trends. *IEEE Wireless Communications*, 17(2), Apr. 2010.

[32] S. Geirhofer, L. Tong, and B. M. Sadler. A cognitive framework for improving coexistence among heterogeneous wireless networks. In *Proc. IEEE GLOBECOM'08*, Nov. 2008.

[33] J. Mitola. Cognitive radio architecture evolution. *Proceedings of the IEEE*, 97(4), Apr. 2009.

[34] R. Chandra, R. Mahajan, T. Moscibroda, R. Raghavendra, and P. Bahl. A case for adapting channel width in wireless networks. In *Proc. ACM SIGCOMM'08*, Aug. 2008.

[35] IEEE 802.22 WG. Ieee p802.22/d0.1 draft standard for wireless regional area networks, part 22: Cognitive wireless ran medium access control (mac) and physical layer (phy) specifications: Policies and procedures for operation in tv bands. IEEE Draft Std., May 2006.

[36] H. Kim and K. G. Shin. Efficient discovery of spectrum opportunities with mac-layer sensing in cognitive radio networks. *IEEE Transactions on Mobile Computing*, 7(5), May 2008.

[37] F. Huang, W. Wang, H. Luo, G. Yu, and Z. Zhang. Prediction-based spectrum aggregation with hardware limitation in cognitive radio networks. In *Proc. IEEE VTC'10-Spring*, May 2010.

[38] P. Bahl, A. Adya, J. Padhye, and A. Wolman. Reconsidering wireless systems with multiple radios. *ACM SIGCOMM Computer Communication Review*, 34(5), Oct. 2004.

[39] D. Xu, E. Jung, and X. Liu. Optimal bandwidth selection in multi-channel cognitive radio networks: How much is too much? In *Proc. IEEE DySPAN'08*, Oct. 2008.

[40] C.-C. Hsu, S. L. Wei, and C.-C. Kuo. A cognitive mac protocol using statistical channel allocation for wireless ad-hoc networks. In *Proc. IEEE WCNC'07*, Mar. 2007.

[41] H. Su and X. Zhang. Cross-layer based opportunistic mac protocols for qos provisioning over cognitive radio wireless networks. *IEEE Journal on Selected Areas in Communications*, 26(1), Jan. 2008.

[42] J. Lee and J. So. Analysis of cognitive radio networks with channel aggregation. In *Proc. IEEE WCNC'10*, Apr. 2010.

[43] X. Zhu, L. Shen, and T.-S. Yum. Analysis of cognitive radio spectrum access with optimal channel reservation. *IEEE Communications Letters*, 11(4), Apr. 2007.

[44] Y. Zhang. Dynamic spectrum access in cognitive radio wireless networks. In *Proc. IEEE ICC'08*, May 2008.

[45] Y. Yuan, P. Bahl, R. Chandra, P. Chou, J. Ferrell, T. Moscibroda, S. Narlanka, and Y. Wu. Knows: Kognitiv networking over white spaces. In *Proc. IEEE DySPAN'07*, Apr. 2007.

[46] Y. Yuan, P. Bahl, R. Chandra, T. Moscibroda, and Y. Wu. Allocating dynamic time-spectrum blocks in cognitive radio networks. In *Proc. ACM MobiHoc'07*, Sept. 2007.

[47] T. Shu and M. Krunz. Throughput-efficient sequential channel sensing and probing in cognitive radio networks under sensing errors. In *Proc. ACM MobiCom'09*, Sept. 2009.

[48] G. Bianchi. Performance analysis of ieee 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3), Mar. 2000.

[49] Huawei. Soft frequency reuse scheme for utran lte. Tech. Report from Huawei, 3GPP R1-050507, May 2005.

[50] WiMAX Forum. Mobile wimax - part i: A technical overview and performance evaluation. Tech. Report from WiMAX Forum, Aug. 2006.

[51] A. L. Stolyar and H. Viswanathan. Self-organizing dynamic fractional frequency reuse in ofdma systems. In *Proc. IEEE INFOCOM'08*, Apr. 2008.

[52] A. L. Stolyar and H. Viswanathan. Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination. In *Proc. IEEE INFOCOM'09*, Apr. 2009.

[53] H. Zhang, Q. Yang, F. Gao, and K. S. Kwak. Distributed adaptive subchannel and power allocation for downlink ofdma with inter-cell interference coordination. In *Proc. IEEE GLOBECOM'10*, Dec. 2010.

[54] V. Tawil, W. Caldwell, I. Reede, T. Kiernan, and C. R. Stevenson. Draft 802.22a (amendment) par. IEEE Std. PAR, IEEE 802.22-09/0029r0, Jan. 2009.

[55] X. Mao, A. Maaref, and K. H. Teo. Adaptive soft frequency reuse for inter-cell interference coordination in sc-fdma based 3gpp lte uplinks. In *Proc. IEEE GLOBECOM'08*, Nov. 2008.

[56] B. Rengarajan, A. L. Stolyar, and H. Viswanathan. Self-organizing dynamic fractional frequency reuse on the uplink of ofdma systems. In *Proc. CISS'10*, Mar. 2010.

[57] F. Wamser, D. Mittelstädt, and D. Staehle. Soft frequency reuse in the uplink of an ofdma network. In *Proc. IEEE VTC'10-Spring*, Jun. 2010.

[58] M. Al-Shalash, F. Khafizov, and Z. Chao. Interference constrained soft frequency reuse for uplink icic in lte networks. In *Proc. IEEE PIMRC'10*, Sept. 2010.

[59] M. Al-Shalash, F. Khafizov, and Z. Chao. Uplink inter-cell interference coordination through soft frequency reuse. In *Proc. IEEE GLOBECOM'10*, Dec. 2010.

[60] H. Kwon and B. G. Lee. Distributed resource allocation through noncooperative game approach in multi-cell ofdma systems. In *Proc. IEEE ICC'06*, Jun. 2006.

[61] Z. Liang, Y. H. Chew, and C. C. Ko. Decentralized bit, subcarrier and power allocation with interference avoidance in multicell ofdma systems using game theoretic approach. In *Proc. IEEE MILCOM'08*, Nov. 2008.

[62] Z. Han, Z. Ji, and K. J. R. Liu. Non-cooperative resource competition game by virtual referee in multi-cell ofdma networks. *IEEE Journal on Selected Areas in Communications*, 25(6), Aug. 2007.

[63] Z. Hou, Y. Cai, and D. Wu. Resource allocation based on integer programming and game theory in uplink multi-cell cooperative ofdma systems. *EURASIP Journal on Wireless Communications and Networking*, Nov. 2011.

[64] S. Buzzi, G. Colavolpe, D. Saturnino, and A. Zappone. Potential games for energy-efficient power control and subcarrier allocation in uplink multicell ofdma systems. *IEEE Journal of Selected Topics in Signal Processing*, 6(2), Apr. 2012.

[65] D. Wu, L. Zhou, Y. Cai, and J. Rodrigues. Energy-efficient resource allocation for uplink ofdma systems using correlated equilibrium. In *Proc. IEEE GLOBECOM'12*, Dec. 2012.

[66] S. Sengupta, R. Chandramouli, S. Brahma, and M. Chatterjee. A game theoretic framework for distributed self-coexistence among ieee 802.22 networks. In *Proc. IEEE GLOBECOM'08*, Dec. 2008.

[67] V. Gardellin, S. K. Das, and L. Lenzini. A fully distributed game theoretic approach to guarantee self-coexistence among wrans. In *Proc. IEEE INFOCOM'10*, May 2010.

[68] V. Gardellin, L. Lenzini, and V. Fontana. Aware channel assignment algorithm for cognitive networks. In *Proc. ACM CoRoNET'11*, Sept. 2011.

[69] A. B. MacKenzie and L. A. DaSilva. *Game Theory for Wireless Engineers*. Morgan & Claypool Publishers, 2006.

[70] H. Kim and K. G. Shin. Asymmetry-aware real-time distributed joint resource allocation in ieee 802.22 wrans. In *Proc. IEEE INFOCOM'10*, Mar. 2010.

[71] J. R. Leigh. *Essentials of Nonlinear Control Theory*. Peter Peregrinus Ltd., 1983.

[72] X. Zhou and H. Zheng. Trust: A general framework for truthful double spectrum auctions. In *Proc. IEEE INFOCOM'09*, Apr. 2009.

[73] J. Jia, Q. Zhang, Q. Zhang, and M. Liu. Revenue generation for truthful spectrum auction in dynamic spectrum access. In *Proc. ACM MobiHoc'09*, May 2009.

[74] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy. A scalable collusion-resistant multi-winner cognitive spectrum auction game. *IEEE Transactions on Communications*, 57(12), Dec. 2009.

[75] S. Sodagari, A. Attar, and S. G. Bilén. On a truthful mechanism for expiring spectrum sharing in cognitive radio networks. *IEEE Journal on Selected Areas in Communications*, 29(4), Apr. 2011.

[76] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, and S. Suri. A general framework for wireless spectrum auctions. In *Proc. IEEE DySPAN'07*, Apr. 2007.

[77] G. S. Kasbekar and S. Sarkar. Spectrum auction framework for access allocation in cognitive radio networks. In *Proc. ACM MobiHoc'09*, May 2009.

[78] L. Chen, S. Iellamo, M. Coupechoux, and P. Godlewski. An auction framework for spectrum allocation with interference constraint in cognitive radio networks. In *Proc. IEEE INFOCOM'10*, Mar. 2010.

[79] X. Wang, Z. Li, P. Xu, Y. Xu, X. Gao, and H.-H. Chen. Spectrum sharing in cognitive radio networks—an auction-based approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(3), Jun. 2010.

[80] J. Jia and Q. Zhang. Competitions and dynamics of duopoly wireless service providers in dynamic spectrum market. In *Proc. ACM MobiHoc'08*, May 2008.

[81] S. Sengupta and M. Chatterjee. An economic framework for dynamic spectrum access and service pricing. *IEEE/ACM Transactions on Networking*, 17(4), Aug. 2009.

[82] L. Duan, J. Huang, and B. Shou. Competition with dynamic spectrum leasing. In *Proc. IEEE DySPAN'10*, Apr. 2010.

[83] K. Zhu, D. Niyato, P. Wang, and Z. Han. Dynamic spectrum leasing and service selection in spectrum secondary market of cognitive radio networks. *IEEE Transactions on Wireless Communications*, 11(3), Mar. 2012.

[84] B. F. Lo. A survey of common control channel design in cognitive radio networks. *Physical Communication*, 4(1), Mar. 2011.

[85] Wikipedia. Vickrey-clarke-groves auction. available online: *http* : *//en.wikipedia.org/wiki/Vickrey − Clarke − Groves_auction.*

[86] A. W. Min, K.-H. Kim, J. P. Singh, and K. G. Shin. Opportunistic spectrum access for mobile cognitive radios. In *Proc. IEEE INFOCOM'11*, Apr. 2011.

[87] Q. Ni and C. C. Zarakovitis. Nash bargaining game theoretic scheduling for joint channel and power allocation in cognitive radio systems. *IEEE Journal on Selected Areas in Communications*, 30(1), Jan. 2012.

[88] Y. Tachwali, B. F. Lo, I. F. Akyildiz, and R. Agustí. Multiuser resource allocation optimization using bandwidth-power product in cognitive radio networks. *IEEE Journal on Selected Areas in Communications*, 31(3), Mar. 2013.

[89] Y. Xu, J. Winter, and W.-C. Lee. Dual prediction-based reporting for object tracking sensor networks. In *Proc. ACM MobiQuitous'04*, Aug. 2004.

[90] Q. Xu, S. Mehrotra, Z. M. Mao, and J. Li. Proteus: Network performance forecast for real-time, interactive mobile applications. In *Proc. ACM MobiSys'13*, Jun. 2013.

[91] W. Burgard, D. Fox, D. Hennig, and T. Schmidt. Estimating the absolute position of a mobile robot using position probability grids. In *Proc. AAAI'96*, Aug. 1996.

[92] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte carlo localization for mobile robots. In *Proc. IEEE ICRA'99*, May 1999.

[93] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2), Feb. 2002.

[94] M. Zhao and W. Wang. A novel semi-markov smooth mobility model for mobile ad hoc networks. In *Proc. IEEE GLOBECOM'06*, Nov. 2006.

[95] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24). online: http://crawdad.org/epfl/mobility/, Feb. 2009.

[96] S. Dimatteo, P. Hui, B. Han, and V. O.K. Li. Cellular traffic offloading through wifi networks. In *Proc. IEEE MASS'11*, Oct. 2011.

[97] X. Ying, J. Zhang, L. Yan, G. Zhang, M. Chen, and R. Chandra. Exploring indoor white spaces in metropolises. In *Proc. ACM MobiCom'13*, Sept. 2013.

[98] H. B. Yilmaz, T. Tugcu, F. Alagöz, and S. Bayhan. Radio environment map as enabler for practical cognitive radio networks. *IEEE Communications Magazine*, 51(12), Dec. 2013.

[99] J.-S. Lee and B. Hoh. Dynamic pricing incentive for participatory sensing. *Pervasive and Mobile Computing*, 6(6), Dec. 2010.

[100] T. Luo, H.-P. Tan, and L. Xia. Profit-maximizing incentive for participatory sensing. In *Proc. IEEE INFOCOM'14*, Apr. 2014.

[101] R. K. Ganti, F. Ye, and H. Lei. Mobile crowdsensing: Current state and future challenges. *IEEE Communications Magazine*, 49(11), Nov. 2011.

[102] D. Yang, G. Xue, X. Fang, and J. Tang. Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing. In *Proc. ACM MobiCom'12*, Aug. 2012.

[103] W. Elmaghraby and P. Keskinocak. Dynamic pricing in the presence of inventory considerations: Research overview, current practices and future directions. *Management Science*, 49(10), Oct. 2003.

[104] M. Cho, M. Fan, and Y.-P. Zhou. Strategic consumer response to dynamic pricing of perishable products. In *Consumer-Driven Demand and Operations Management Models*. Springer US, 2009.

[105] S. Dasu and C. Tong. Dynamic pricing when consumers are strategic: Analysis of posted and contingent pricing schemes. *European Journal of Operational Research*, 204(3), Aug. 2010.