K. J. Ray Liu and
Beibei Wang

Cognitive Radio Networking and Security

A Game-Theoretic View

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521762311

This page intentionally left blank

Cognitive Radio Networking and Security

With the rapid growth of new wireless devices and applications over the past decade, the demand for wireless radio spectrum is increasing relentlessly. The development of cognitive radio networking provides a framework for making the best possible use of limited spectrum resources, and it is revolutionizing the telecommunications industry.

This book presents the fundamentals of designing, implementing, and deploying cognitive radio communication and networking systems. Uniquely, it focuses on game theory and its applications to various aspects of cognitive networking. It covers in detail the core aspects of cognitive radio, including cooperation, situational awareness, learning, and security mechanisms and strategies. In addition, it provides novel, state-of-the-art concepts and recent results. This is an ideal reference for researchers, students, and professionals in industry who need to learn the applications of game theory to cognitive networking.

K. J. Ray Liu is a Distinguished Scholar-Teacher at the University of Maryland, College Park. He is the recipient of numerous honors and awards including the 2009 IEEE Signal Processing Society Technical Achievement Award, IEEE Signal Processing Society Distinguished Lecturer, National Science Foundation Presidential Young Investigator, and various best-paper awards.

Beibei Wang is currently a Senior Systems Engineer with Corporate Research and Development, Qualcomm Incorporated. She received her Ph.D. from the University of Maryland, College Park in 2009. Her research interests include dynamic spectrum allocation and management in cognitive radio systems, cooperative communications, multimedia communications, game theory and learning, and network security.

Cognitive Radio Networking and Security

A Game-Theoretic View

K. J. RAY LIU
University of Maryland, College Park

BEIBEI WANG
Qualcomm Incorporated

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521762311

First published in print format 2010

ISBN-13 978-0-511-90418-9 eBook (Adobe Reader)

ISBN-13 978-0-521-76231-1 Hardback

In memory of my great-grand mother
Lang-Xiang Liu (Kane Koda), August 4, 1899–April 11, 1992, for the eternal loving bond transcending generations.
I always miss you. – K. J. Ray Liu

To my parents, Liangyuan Wang and Shuqin Huang, for their unconditional love and support. – Beibei Wang

Contents

Preface

Recent increases in demand for cognitive radio technology have driven researchers and technologists to rethink the implications of the traditional engineering designs and approaches to communications and networking. One issue is that the traditional thinking is that one should try to have more bandwidth, more resources, and more of everything, while we have come to the realization that the problem is not that we do not have enough bandwidth or resources. It is rather that the bandwidth/resource utilization rates in many cases are too low. For example, the TV bandwidth utilization nowadays in the USA is less than 6%, which is quite similar to that in most developed countries. So why continue wanting to obtain more new bandwidth when it is indeed a scarce commodity already? Why not just utilize the wasted resource in a more effective way?

Another reconsideration is that often one can find the optimization tools and solutions employed in engineering problems being too rigid, without offering much flexibility, adaptation, and learning. The super highway is a typical example in that, during traffic hours, one direction is completely jammed with bumper-to-bumper cars, while the other direction has few cars with mostly empty four-lane way. That is almost the case for networking as well. Rigid, inflexible protocols and strategies often leave wasted resources that could otherwise be efficiently utilized by others. It was recognized that traditional communication and networking paradigms have taken little or no situational information into consideration by offering cognitive processing, reasoning, learning, and adaptation. Along the same lines, such awareness also drives us to seek an optimization tool to better enhance cooperation and resolve conflict with learning capability.

In the past decade we have witnessed that the concept of cognitive networking and communications has offered a revolutionary perspective in the design of modern communication infrastructure. By cognitive communications and networking we mean that a communication system is composed of elements that can dynamically adapt themselves to the varying conditions, resources, environments, and users through interacting, learning, and reasoning to evolve and reach better operating points or a better set of system parameters to enhance cooperation and resolve conflict, if any. Those factors can include awareness of channel conditions, energy efficiency, bandwidth availability, locations, spectrum usage, and the connectivity of a network, to name just a few. Such design with awareness of situations, resources, environments, and users forms the core concept of the emerging field of cognitive communications and networking. Many new ideas have thus been inspired and have blossomed.

Cognitive radio, a special case of cognitive networking, has received a lot of attention recently. In contrast to traditional radio, cognitive radio is an intelligent wireless communication system that is aware of its surrounding environment and can adaptively change its operating parameters on the basis of interactions with the environment and users. With cognitive radio technology, future wireless devices are envisioned to be able to sense and analyze their surrounding environment and user conditions, learn from the environmental variations, and adapt their operating parameters to achieve highly reliable communications and efficient utilization of the spectrum resources.

In a cognitive network, nodes are intelligent and have the ability to observe, learn, and act to optimize their performance. Since nodes generally belong to different authorities and pursue different goals, fully cooperative behaviors cannot be taken for granted. Instead, nodes will cooperate with others only when cooperation can improve their own performance. Often nodes with such selfish behaviors are regarded as rational. Therefore, a key problem in cognitive networks is how to stimulate cooperation among selfish nodes. To address the interactions of the dynamics among conditions, resources, environments, and users, game theory has naturally become an important emerging tool that is ideal and essential in studying, modeling, and analyzing the cognitive interaction process. This is especially true because a rational user in a cognitive network often behaves selfishly to maximize his/her own utility or welfare. There is of course no surprise here, since game theory has been a core tool in the study of economics and business/social models, in particular in the understanding of cooperation, interaction, and conflict, via which strategies and mechanisms can be developed to offer flexible and adaptable solutions.

In recent years, it has found a major engineering challenge in the emerging development of cognitive communications and networking. In a certain sense, what is taking place in cognitive communications and networking can be viewed as a kind of information game, where optimal policies, strategies, and protocols are developed from the signals/information obtained by users through interaction, cooperation, or competition of communication/networking devices, rather than economic and financial games being played in human society. Not only can traditional games be leveraged to apply to various networking scenarios, but also new games can be developed, since wireless communication is interference-limited instead of quantity-limited as is the case for most economic models. Therefore we are seeing the new era of information games emerging and unfolding.

This book aims at providing a comprehensive coverage of fundamental issues on cooperation, learning, adaption, and security that should be understood in the design, implementation, and deployment of cognitive communication and networking systems, with a focus on game-theoretical approaches. Most of the material stems from our research over the past decade pursuing the realization of cognitive communications and secure networking. A goal of the book is to provide a bridge between advanced research on the one hand and classroom learning and self-study on the other by offering an emphasis on systematic game-theoretical treatments of cognitive communications and networking. In particular, we partition the book into three parts.

In Part I, we address the issues relating to cognitive radio communications and user cooperation. The users in a cognitive network will be assumed to be rational when cooperating with others, i.e., they behave selfishly in maximizing their own interest. In Chapter 1 we provide an introductory overview and survey of cognitive radio technology and related technical issues, including spectrum sensing, dynamic spectrum sharing and allocation, and cognitive radio platforms and standards, followed by a tutorial on fundamentals of game theory for cognitive networking in Chapter 2. We then focus on each important component of cognitive radio technology with more detailed treatments. Chapter 3 introduces Markov models for efficient dynamic spectrum allocation. Chapter 4 considers repeated open spectrum sharing games with cheat-proof strategies. The concept of pricing games is studied in Chapter 5 for dynamic spectrum allocation. A multi-winner spectrum auction game is presented in Chapter 6 to address the interference-limited situation of wireless communications. An evolutionary cooperative spectrum sensing game is then introduced in Chapter 7 in order for the reader to understand the best strategy for cooperation and its evolution when the situation is changing. It is followed by discussion of a stochastic anti-jamming game to design the optimal adaptive defense strategies against cognitive malicious attackers in Chapter 8. Finally, the issue of opportunistic multiple access for cognitive networks with cooperation of relays is studied in Chapter 9.

In Part II, the focus is on resource awareness and learning. The discussion is extended beyond the narrow definition of a cognitive radio to the general notion of cognitive wireless communications and networking. Various situational awareness and learning scenarios are considered. In Chapter 10, reinforcement learning for energy awareness is discussed. Chapter 11 considers a repeated game framework and learning for cooperation enforcement. Dynamic pricing games for routing are studied in Chapter 12. A graph-theoretical connectivity-aware approach for network lifetime optimization is presented in Chapter 13, followed by the issues relating to graph-theoretic network maintenance and repair in Chapter 14.

Because of the interactions and cooperation in cognitive networks, security becomes a major issue. Therefore Part III is dedicated to the consideration of a securing mechanism and strategies. However, since there is no consensus notion of a security paradigm yet in this arena, there are three main themes in this part: trust modeling and evaluation, defense mechanisms and strategies, and game-theoretical analysis of security. Some users who are attackers are assumed to be malicious, i.e., their goal is to damage the system's performance, instead of maximizing their own interest. Since security in centralized systems is less of an issue, most of the chapters are formulated in terms of distributed ad hoc networking. First information-theoretical trust models and an evaluation framework are presented in Chapter 15 for network security, followed by some defenses against a series of attacks such as routing disruption attacks in Chapter 16 and injecting traffic attacks in Chapter 17. Attack-resistant mechanisms and optimal strategies for cooperation stimulation are considered in Chapters 18 and 19, respectively. Finally, statistical securing approaches for cooperation stimulation and enforcement under noise and imperfect monitoring situations are presented in the next three chapters, with Chapter 20 focusing on belief evaluation and vulnerability analysis,

Chapter 21 on defense against insider attacks, and Chapter 22 on secure cooperation stimulation.

This book is intended to be a textbook or a reference book for graduate-level courses on wireless communications and networking that cover cognitive radios, game theory, and/or security. We hope that the comprehensive coverage of cognitive communications, networking, and security with a holistic treatment from the view of information games will make this book a useful resource for readers who want to understand this emerging technology, as well as for those who are conducting research and development in this field.

This book could not have been made possible without the research contributions by the following people: Charles Clancy, Amr El-Sherif, Zhu Han, Ahmed Ibrahim, Zhu Ji, Charles Pandana, Karim Seddik, Yan Sun, Yongle Wu, and Wei Yu. We also would like to thank all the colleagues whose work enlightening our thoughts and research made this book possible. We can only stand on the shoulders of giants.

K. J. Ray Liu
Beibei Wang

Part I

Cognitive radio communications and cooperation

1 Introduction to cognitive radios

With the rapid deployment of new wireless devices and applications, the last decade has witnessed a growing demand for wireless radio spectrum. However, the policy of fixed spectrum assignment produces a bottleneck for more efficient spectrum utilization, such that a great portion of the licensed spectrum is severely under-utilized. The inefficient usage of the limited spectrum resources has motivated the regulatory bodies to review their policy and start to seek innovative communication technology that can exploit the wireless spectrum in a more intelligent and flexible way. The concept of cognitive radio was proposed to address the issue of spectrum efficiency and has been receiving increasing attention in recent years, since it equips wireless users with the capability to optimally adapt their operating parameters according to the interactions with the surrounding radio environment. There have been many significant developments in the past few years concerning cognitive radios. In this chapter, the fundamentals of cognitive radio technology, including the architecture of a cognitive radio network and its applications, are introduced. The existing works on spectrum sensing are reviewed, and important issues in dynamic spectrum allocation and sharing are discussed in detail. Finally, an overview on implementation of cognitive radio platforms and standards for cognitive radio technology is provided.

1.1 Introduction

The usage of radio spectrum resources and the regulation of radio emissions are coordinated by national regulatory bodies such as the Federal Communications Commission (FCC). The FCC assigns spectrum to licensed holders, also known as *primary users*, on a long-term basis for large geographical regions. However, a large portion of the assigned spectrum remains under-utilized as illustrated in Figure 1.1 [114]. The inefficient usage of the limited spectrum necessitates the development of dynamic spectrum access techniques, where users who have no spectrum licenses, also known as *secondary users*, are allowed to use the temporarily unused licensed spectrum. In recent years, the FCC has been considering more flexible and comprehensive uses of the available spectrum [116], through the use of *cognitive radio* technology [284].

Cognitive radio is the key enabling technology that enables next-generation (xG) communication networks, also known as dynamic spectrum access (DSA) networks, to utilize the spectrum more efficiently in an opportunistic fashion without interfering

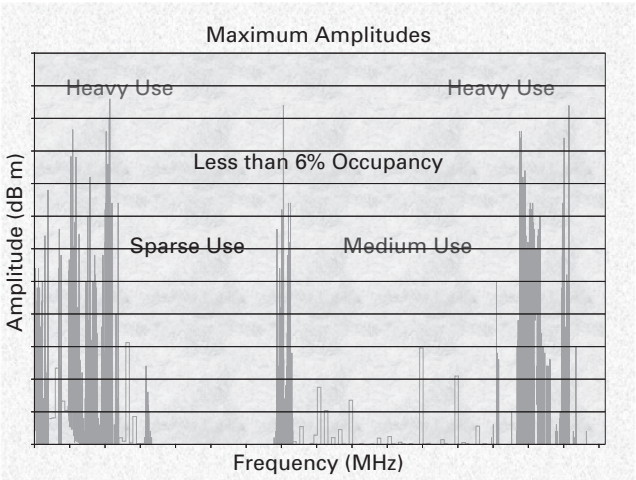

Figure 1.1 Spectrum usage, from FCC Report [7].

with the primary users. It is defined as a radio that can change its transmitter parameters according to the interactions with the environment in which it operates [114]. It differs from conventional radio devices in that a cognitive radio can equip users with *cognitive capability* and *reconfigurability* [160] [7]. Cognitive capability refers to the ability to sense and gather information from the surrounding environment, such as information about the transmission frequency, bandwidth, power, modulation, etc. With this capability, secondary users can identify the best available spectrum. Reconfigurability refers to the ability to rapidly adapt the operational parameters according to the sensed information in order to achieve the optimal performance. By exploiting the spectrum in an opportunistic fashion, cognitive radio enables secondary users to sense which portions of the spectrum are available, select the best available channel, coordinate spectrum access with other users, and vacate the channel when a primary user reclaims the spectrum-usage right.

Considering the more flexible and comprehensive use of the spectrum resources, especially when secondary users coexist with primary users, traditional spectrum-allocation schemes and spectrum-access protocols are no longer applicable. New approaches to spectrum management need to be developed to solve new challenges in research related to cognitive radio, specifically in spectrum sensing and dynamic spectrum sharing.

Since primary users have priority in using the spectrum, when secondary users coexist with primary users, they have to perform real-time wideband monitoring of the licensed spectrum to be used. When secondary users are allowed to transmit data simultaneously with a primary user, the interference temperature limit should not be violated [65]. If secondary users are allowed to transmit only when the primary users are not using the spectrum, they need to be aware of the primary users' reappearance through various detection techniques, such as energy detection, feature detection, matched filtering, and coherent detection. Owing to noise uncertainty, shadowing, and multipath effects, the

detection performance of single-user sensing is pretty limited. Cooperative sensing has been considered effective in improving detection accuracy by taking advantage of the spatial and multiuser diversity. In cooperative spectrum sensing, how to select proper users for sensing, how to fuse an individual user's decision and exchange information, and how to perform distributed spectrum sensing are issues worth studying.

In order to fully utilize the spectrum resources, efficient dynamic spectrum allocation and sharing schemes are very important. Novel spectrum-access control protocols and control-channel management should be designed to accommodate the dynamic spectrum environment while avoiding collision with a primary user. When a primary user reappears in a licensed band, a good spectrum-handoff mechanism is required to provide secondary users with smooth frequency transition with low latency. In multi-hop cognitive wireless networks, intermediate cognitive nodes should intelligently support relaying information and routing through using a set of dynamically changing channels. In order to manage the interference to the primary users and the mutual interference among themselves, secondary users' transmission power should be carefully controlled, and their competition for the spectrum resources should also be addressed.

There have been many significant developments relating to cognitive radios in the past few years. In Section 1.2, we overview the fundamentals of cognitive radio technology, including the architecture of a cognitive radio network and its applications. In Section 1.3, we review existing works on spectrum sensing, including interference temperature, different types of detection techniques, and cooperative spectrum sensing. In Section 1.4 we discuss several important issues in dynamic spectrum allocation and sharing. Finally, we present in Section 1.5 several cognitive radio platforms that have been developed in research institutes and industry, and standards on cognitive radio technology.

1.2 Fundamentals

1.2.1 Cognitive radio characteristics

The dramatic increase of service quality and channel capacity in wireless networks is severely limited by the scarcity of energy and bandwidth, which are the two fundamental resources for communications. Therefore, researchers are currently focusing their attention on new communications and networking paradigms that can intelligently and efficiently utilize these scarce resources. Cognitive radio (CR) is one critical enabling technology for future communications and networking that can utilize the limited network resources in a more efficient and flexible way. It differs from traditional communication paradigms in that the radios/devices can adapt their operating parameters, such as transmission power, frequency, modulation type, etc., to the variations of the surrounding radio environment [114]. Before CRs adjust their operating mode to environment variations, they must first gain necessary information from the radio environment. This kind of characteristic is referred to as *cognitive capability* [160], which enables CR devices to be aware of the transmitted

waveform, radio-frequency (RF) spectrum, communication-network type/protocol, geographical information, locally available resources and services, user needs, security policy, and so on. After CR devices have gathered the needed information from the radio environment, they can dynamically change their transmission parameters according to the sensed environment variations and achieve optimal performance, which is referred to as *reconfigurability* [160]. For instance, the frequencies of available spectrum bands may keep changing, due to primary users' transmission. Secondary users equipped with CR will know which portion of the spectrum is not occupied by sensing the spectrum, and tune their transmitting frequencies to the spectrum white space.

1.2.2 Cognitive radio functions

A typical duty cycle of CR, as illustrated in Figure 1.2, includes detecting spectrum white space, selecting the best frequency bands, coordinating spectrum access with other users, and vacating the frequency when a primary user appears. Such a cognitive cycle is supported by the following functions:

- *spectrum sensing and analysis;*
- *spectrum management and handoff;*
- *spectrum allocation and sharing.*

Through spectrum sensing and analysis, CR can detect the spectrum white space (see Figure 1.3), i.e., a portion of the frequency band that is not being used by the primary users, and utilize the spectrum. On the other hand, when primary users start using the licensed spectrum again, CR can detect their activity through sensing, so that no harmful interference is generated due to secondary users' transmission.

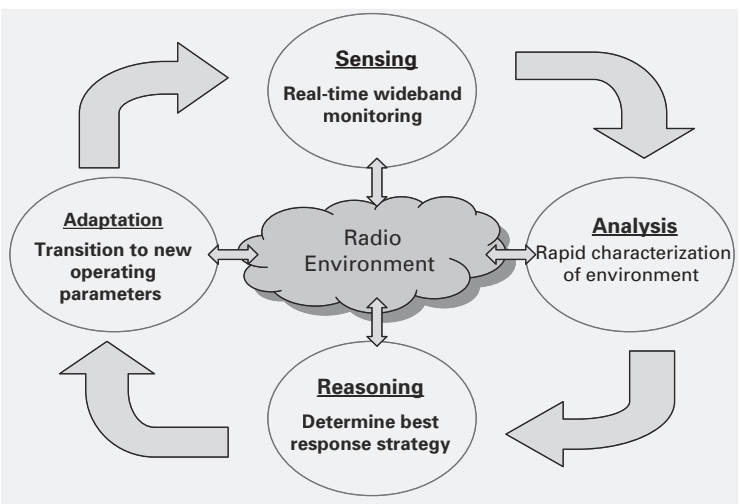

Figure 1.2 The cognitive cycle.

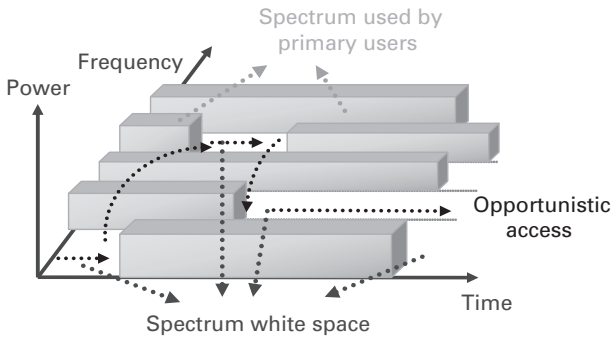

Figure 1.3 Spectrum white space.

After recognizing the spectrum white space by sensing, the spectrum management and handoff function of CR enables secondary users to choose the best frequency band and hop among multiple bands according to the time-varying channel characteristics to meet various quality-of-service (QoS) requirements [7]. For instance, when a primary user reclaims his/her frequency band, the secondary user using the licensed band can direct his/her transmission to other available frequencies, according to the channel capacity determined by the noise and interference levels, path loss, channel error rate, holding time, etc.

In dynamic spectrum access, a secondary user may share the spectrum resources with primary users, other secondary users, or both. Hence, a good mechanism for spectrum allocation and sharing is critical in order to achieve high spectrum efficiency. Since primary users own the spectrum rights, when secondary users coexist in a licensed band with primary users, the interference level due to secondary spectrum usage should be limited by a certain threshold. When multiple secondary users share a frequency band, their access should be coordinated in order to alleviate collisions and interference.

1.2.3 Network architecture and applications

With the development of CR technologies, secondary users who have not been allocated spectrum-usage rights can utilize the temporally unused licensed bands owned by the primary users. Therefore, in a CR network architecture, the components include both a secondary network and a primary network, as shown in Figure 1.4.

A secondary network is a network composed of a set of secondary users and one or more secondary base stations. Secondary users can access the licensed spectrum only when it is not occupied by a primary user. The opportunistic spectrum access of secondary users is usually coordinated by a secondary base station, which is a fixed infrastructure component serving as a hub of the secondary network. Both secondary users and secondary base stations are equipped with CR functions. If several secondary networks share one common spectrum band, their spectrum usage may be coordinated by a central network entity, called a *spectrum broker* [372]. The spectrum broker collects operation information from each secondary network, and allocates the network resources in such a way as to achieve efficient and fair spectrum sharing.

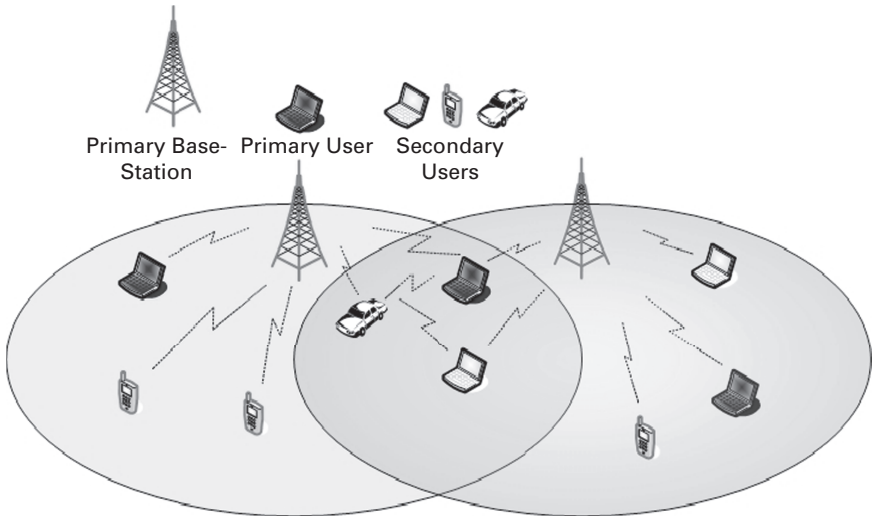

Figure 1.4 Dynamic spectrum sharing.

A primary network is composed of a set of primary users and one or more primary base stations. Primary users are authorized to use certain licensed spectrum bands under the coordination of primary base stations. Their transmission should not be interfered with by secondary networks. Primary users and primary base stations are in general not equipped with CR functions. Therefore, if a secondary network shares a licensed spectrum band with a primary network, besides detecting the spectrum white space and utilizing the best spectrum band, the secondary network is required to immediately detect the presence of a primary user and direct the secondary transmission to another available band so as to avoid interfering with primary transmission.

Because CRs are able to sense, detect, and monitor the surrounding RF environment such as interference and access availability, and reconfigure their own operating characteristics to best match outside situations, cognitive communications can increase spectrum efficiency and support higher-bandwidth service. Moreover, the capability of real-time autonomous decisions for efficient spectrum sharing also reduces the burdens of centralized spectrum management. As a result, CRs can be employed in many applications.

First, the capacity of military communications is limited by radio spectrum scarcity because static frequency assignments freeze bandwidth into unproductive applications, where a large amount of spectrum is idle. CR using dynamic spectrum access can alleviate the spectrum congestion through efficient allocation of bandwidth and flexible spectrum access [284]. Therefore, CR can provide military users with adaptive, seamless, and secure communications.

Moreover, a CR network can also be implemented to enhance public safety and homeland security. A natural disaster or terrorist attack can destroy existing communication infrastructure, so an emergency network to aid the search and rescue effort becomes indispensable. Since a CR can recognize spectrum availability and reconfigure itself for much more efficient communication, this provides public-safety personnel with

dynamic spectrum selectivity and reliable broadband communication to minimize information delay. Moreover, CR can facilitate interoperability between various communication systems. Through adapting to the requirements and conditions of another network, the CR devices can support multiple service types, such as voice, data, video, etc.

Another very promising application of CR is in the commercial markets for wireless technologies. Since CR can intelligently determine which communication channels are in use and automatically switches to an unoccupied channel, it provides additional bandwidth and versatility for rapidly growing data applications. Moreover, CR constantly scans the entire band to look for and avoid interference from other users; whenever a channel in use by CR is reclaimed by a primary user or interfered with by another secondary user, the CR will instantly select a free channel from its constantly updated free-channel list. The adaptive and dynamic channel switching can help avoid spectrum conflict and expensive redeployment. In addition, since CR can utilize a wide range of frequencies, some of which have excellent propagation characteristics, CR devices are less susceptible to fading related to growing foliage, buildings, terrain, and weather. A CR configuration can also support mobile applications with low cost. When frequency changes are needed due to conflict or interference, the CR frequency-management software will change the operating frequency automatically even without human intervention. Additionally, the radio software can change the service bandwidth remotely to accommodate new applications. As long as no end-user hardware needs to be updated, product upgrades or configuration changes can be completed simply by downloading newly released radio management software. Thus, CR is viewed as the key enabling technology for future mobile wireless services anywhere, anytime, and with any device.

1.3 Spectrum sensing and analysis

Through spectrum sensing, CR can obtain necessary observations about its surrounding radio environment, such as the presence of primary users and the appearance of spectrum holes. Only with this information can CR adapt its transmitting and receiving parameters, such as transmission power, frequency, modulation schemes, etc., in order to achieve efficient spectrum utilization. Therefore, spectrum sensing and analysis is the first critical step toward dynamic spectrum management. In this section, we will discuss three different aspects of spectrum sensing. First is the interference temperature model, which measures the interference level observed at a receiver and is used to protect licensed primary users from harmful interference due to unlicensed secondary users. Then we will talk about spectrum hole detection to determine additional available spectrum resources and compare several detection techniques. Finally, we will discuss cooperative sensing with multiple users or relays' help.

1.3.1 Interference temperature

Secondary users do not have a license for using the spectrum, and can use the licensed spectrum only when they cause no harmful interference to primary users. This requires

secondary users to be equipped with CRs, which can detect primary users' appearance and decide which portion of the spectrum is available. Such a decision can be made according to various metrics. The traditional approach is to limit the transmitter power of interfering devices, i.e., the transmitted power should be no more than a prescribed noise floor at a certain distance from the transmitter. However, due to the increased mobility and variability of RF emitters, constraining the transmitter power becomes problematic, since unpredictable new sources of interference may appear. To address this issue, the FCC Spectrum Policy Task Force [115] has proposed a new metric on interference assessment, the *interference temperature*, to enforce an interference limit perceived by receivers. The interference temperature is a measure of the RF power available at a receiving antenna to be delivered to a receiver, reflecting the power generated by other emitters and noise sources [236]. More specifically, it is defined as the temperature equivalent to the RF power available at a receiving antenna per unit bandwidth [66], i.e.,

$$T_{\mathrm{I}}(f_{\mathrm{c}}, B) = \frac{P_{\mathrm{I}}(f_{\mathrm{c}}, B)}{kB}, \tag{1.1}$$

where $P_{\mathrm{I}}(f_{\mathrm{c}}, B)$ is the average interference power in watts centered at f_{c}, covering bandwidth B measured in hertz, and Boltzmann's constant k is $1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$.

With the concept of interference temperature, the FCC further established an *interference-temperature limit*, which provides a maximum amount of tolerable interference for a given frequency band at a particular location. Any unlicensed secondary transmitter using this band must guarantee that their transmission plus the existing noise and interference will not exceed the interference-temperature limit at a licensed receiver.

Since any transmission in the licensed band is viewed to be harmful if it would increase the noise floor above the interference-temperature limit, it is necessary that the receiver have a reliable spectral estimate of the interference temperature. This requirement can be fulfilled by using the multitaper method to estimate the power spectrum of the interference temperature with a large number of sensors [160]. The multitaper method can solve the tradeoff between bias and variance of an estimator and provide a near-optimal estimation performance. The large number of sensors can account for the spatial variation of the RF energy from one location to another. A subspace-based method to gain knowledge of the quality and usage of a spectrum band has also been proposed [454], in which information about the interference temperature is obtained by eigenvalue decomposition.

Given a particular frequency band in which the interference-temperature limit is not exceeded, that band could be made available for secondary usage. If a regulatory body sets an interference-temperature limit T_{L} for a particular frequency band with bandwidth B, then the secondary transmitter has to keep the average interference below kBT_{L}. Therefore, the interference temperature serves as a cap placed on the potential RF energy that could appear on that band, and there have been some previous studies on how to implement efficient spectrum allocation with the interference-temperature limit.

In [66], two interpretations of the interference-temperature models were analyzed, since there is ambiguity over which signals are considered interference, and

which frequency f_c and bandwidth B to use. The first is the *ideal interference-temperature model*, in which interference is limited specifically to primary signals. Assume a secondary transmitter is operating with average power P in a band $[f_c - B/2, f_c + B/2]$, which overlaps n primary signals with frequency f_i and bandwidth B_i. Then, the interference-temperature limit will ensure that

$$T_I(f_i, B_i) + \frac{MP}{kB_i} \leq T_L(f_i), \qquad \forall 1 \leq i \leq n, \tag{1.2}$$

where M represents attenuation due to fading and path loss between the secondary transmitter and the primary receiver. However, it is generally very difficult to distinguish primary signals from secondary signals or measure T_I in the presence of a primary signal, unless some a priori information is known about the primary signal. Therefore, a *generalized model* is considered, which requires no a priori knowledge of the RF environment and limits the secondary transmitter's parameters, since the information about the primary receivers is unknown. In the generalized model, the interference-temperature limit is applied to the entire frequency range, i.e.,

$$T_I(f_c, B) + \frac{MP}{kB} \leq T_L(f_c). \tag{1.3}$$

With the interference-temperature-limit constraints, secondary users can select the optimal operating frequency, bandwidth, and power to maximize their capacity. Spectrum shaping has been proposed to improve spectrum efficiency [84] in CR networks. More specifically, using interference fitting, a CR senses the shape of the interference power spectrum and creates spectra inversely shaped with respect to the current interference environment in order to take advantage of gaps between the noise floor and the cap of the interference-temperature limit. Another application of spectrum shaping is to create notched power spectra that allow the usage of noncontiguous spectrum segments and avoid primary signals. Dynamic spectrum access with QoS and interference-temperature constraints has been studied in [472]. The objective of the scheme is to maximize the total throughput of all secondary users in a network, constrained by a minimum QoS requirement and a total-received-power requirement at a specified measurement point. Within the framework of the interference-temperature model, the work in [399] presented cooperative algorithms for selecting the most appropriate channel for transmission in a cognitive mesh network. Each mesh node computes a set of channels available for transmission without violating the interference-temperature limit in its interference range, and then uses a per-hop link-cost metric and an end-to-end routing metric to select channels for each hop on the path.

Traditional interference constraints are usually binary and inefficient, since they consider only pair-wise sets of users. Non-binary constraints in line with the interference-temperature model have been studied in [39], which considered the effects of multiple interference sources from across the network. Under these constraints, simultaneous spectrum assignment for a number of secondary transmitters is achievable to improve spectrum utilization, while ensuring that the primary receivers can maintain

a certain QoS. The interference-temperature dynamics in a CR network were investigated in [400] using a hidden Markov model (HMM). The HMM is trained with the observed interference-temperature values using the Baum–Welch procedure. The trained HMM is shown to be statistically stable, and can be used as a sequence generator for secondary nodes to predict the interference temperature of the channel in the future. Secondary users can further utilize the prediction to aid their channel selection for transmission.

A comprehensive analysis has been presented in [67], which quantifies how interference-temperature limits should be selected and how those choices affect the range of licensed signals. Assuming a fixed transmit bandwidth that overlaps a single primary signal, [67] first determines the base interference temperature seen by a set of network nodes, and then quantifies the total network capacity achievable by the underlay secondary network. It is shown that the capacity achieved is a simple function of the number of nodes, the average bandwidth, and the fractional impact on the primary signal's coverage area. On the basis of the capacity analysis, [67] introduces a PHY/MAC protocol, interference-temperature multiple access (ITMA), by first sensing the RF environment and then determining the bandwidth and power needed in order to achieve a desired capacity. However, as observed in [67], the capacity achievable from the interference-temperature model is low, compared with the amount of interference with primary users it can cause. It is also argued by other commenting parties of the FCC that the interference-temperature approach is not a workable concept and would result in increased interference in the frequency bands where it would be used. Therefore, in May 2007 the FCC terminated the work on rule making for implementing the interference-temperature model.

1.3.2 Spectrum sensing

Spectrum sensing enables the capability of a CR to measure, learn, and be aware of the radio's operating environment, such as the spectrum availability and interference status. When a certain frequency band is detected as not being used by the primary licensed user of the band at a particular time in a particular position, secondary users can utilize the spectrum, i.e., there exists a spectrum opportunity. Therefore, spectrum sensing can be performed in the time, frequency, and spatial domains. With the recent development of beamforming technology, multiple users can utilize the same channel/frequency at the same time in the same geographical location. Thus, if a primary user does not transmit in all directions, extra spectrum opportunities can be created for secondary users in the directions where the primary user is not operating, and spectrum sensing needs also to take the angles of arrivals (AoAs) into account [270] [100] [477]. Primary users can also use their assigned bands by means of spread spectrum or frequency hopping, and then secondary users can transmit in the same band simultaneously without severely interfering with primary users as long as they adopt an orthogonal code with respect to the codes adopted by primary users [180] [426]. This creates spectrum opportunities in the code domain, but meanwhile requires detection of the codes used by primary users as well as multipath parameters.

Table 1.1. Summary of main spectrum-sensing techniques

Type	Test statistics	Advantages	Disadvantages
Energy detector	Energy of the received signal samples	• Easy to implement • Does not require prior knowledge about primary signals	• High false-alarm rate due to noise uncertainty • Very unreliable in low-SNR regimes • Cannot differentiate a primary user from other signal sources
Feature detector	Cyclic spectrum density function of the received signal, or by matching general features of the received signal to the already-known primary-signal characteristics	• More robust against noise uncertainty and better detection in low-SNR regimes than energy detection • Can distinguish among different types of transmissions and primary systems	• Specific features, e.g., cyclostationary features, must be associated with primary signals • Particular features may need to be introduced, e.g., to OFDM-based communications
Matched filtering and coherent detection	Projected received signal in the direction of the already-known primary signal or a certain waveform pattern	• More robust against noise uncertainty and better detection in low-SNR regimes than feature detector • Require fewer signal samples to achieve good detection	• Require precise prior information about certain waveform patterns of primary signals • High complexity

A wealth of literature on spectrum sensing focuses on primary transmitter detection based on the local measurements of secondary users, since detecting the primary users that are receiving data is in general very difficult. According to the a priori information they require and the resulting complexity and accuracy, spectrum-sensing techniques can be categorized into the following types, which are summarized in Table 1.1.

1.3.2.1 Energy detector

Energy detection is the most common type of spectrum sensing because it is easy to implement and requires no a priori knowledge about the primary signal.

Assume the hypothesis model of the received signal is

$$\begin{aligned} \mathcal{H}_0 &: y(t) = n(t), \\ \mathcal{H}_1 &: y(t) = hx(t) + n(t), \end{aligned} \tag{1.4}$$

where $x(t)$ is the primary user's signal to be detected at the local receiver of a secondary user, $n(t)$ is the additive white Gaussian noise (AWGN), and h is the channel gain from the primary user's transmitter to the secondary user's receiver. $\mathcal{H}_0$ is a null hypothesis, meaning that there is no primary user present in the band, while $\mathcal{H}_1$ means the primary user's presence. The detection statistics of the energy detector can be defined as the average (or total) energy of N observed samples

$$T = \frac{1}{N}\sum_{t=1}^{N}|y(t)|^2. \tag{1.5}$$

The decision on whether the spectrum is being occupied by the primary user is made by comparing the detection statistics T with a predetermined threshold λ. The performance of the detector is characterized by two probabilities: the probability of a false alarm P_{F} and the probability of detection P_{D}. P_{F} denotes the probability that the hypothesis test decides $\mathcal{H}_1$ while it is actually $\mathcal{H}_0$, i.e.,

$$P_{\mathrm{F}} = \Pr(T > \lambda|\mathcal{H}_0). \tag{1.6}$$

P_{D} denotes the probability that the test correctly decides $\mathcal{H}_1$, i.e.,

$$P_{\mathrm{D}} = \Pr(T > \lambda|\mathcal{H}_1). \tag{1.7}$$

A good detector should ensure a high detection probability P_{D} and a low false-alarm probability P_{F}, or it should optimize the spectrum-usage efficiency (e.g., the QoS of a secondary-user network) while guaranteeing a certain level of primary-user protection. To this end, various approaches to improve the efficiency of energy-detector-based spectrum sensing have been proposed.

Since the detection performance is very sensitive to the noise-power-estimate error [422], an adaptive noise-level-estimation approach is proposed in [316], where a multiple-signal classification algorithm is used to decouple the noise and signal subspaces and estimate the noise floor. A constant false-alarm-rate threshold is further computed to study the spectrum occupancy and its statistics. A well-chosen detection threshold can minimize spectrum-sensing error, provide the primary user with enough protection, and fully enhance spectrum utilization. In [441] the detection threshold is optimized iteratively to satisfy the requirement on the false-alarm probability. Threshold optimization subject to spectrum-sensing constraints is investigated in [320], where an optimal adaptive threshold level is developed by utilizing the spectrum-sensing error function. In [259], forward methods for energy detection are proposed, for which the noise power is unknown and is adaptively estimated. In order to find and localize narrowband signals, a localization algorithm based on double-thresholding (LAD) is proposed in [437], where the usage of two thresholds can provide signal separation and localization. The LAD method involves blind narrowband-signal detection, and neither information about the noise level nor narrowband signals are required. The LAD method with normalized thresholds can reduce computational complexity without performance loss, and the estimation of the number of narrowband signals is made more accurate by combining adjacent clusters. The sensing-throughput tradeoff of energy detection is studied in [265], where the duration of the sensing period in a time slot is optimized to maximize the achievable throughput for the secondary users under the constraint that the primary users are sufficiently protected. A novel wideband spectrum-sensing technique based on energy detection is introduced in [355], which jointly detects the signal energy levels over multiple frequency bands in order to improve the opportunistic throughput of CRs and reduce their interference with the primary systems. The analysis in [443] shows that detection of narrowband transmission using energy

detection over multi-band OFDM is feasible, and can be further extended to cover more complex systems.

Experimental studies on energy-detection-based spectrum access have also been conducted in the literature. [365] proposes spectrum-sensing algorithms for localizing transmitters in the same band by applying triangulation techniques based on sensed power at each sensor, and addresses the issue of how to find the spectral occupancy of multiple transmitters over a wide range of frequencies. In [150], actual measurements of the channel-access pattern in the 2.4-GHz ISM band are taken. This approach uses a vector signal analyzer to collect complex baseband data, and uses the collected data to statistically characterize the idle and busy periods of the channel. The performance of energy detection under various channel conditions is studied in [97], where the probability of detection under conditions of AWGN and fading channels is derived.

Besides its low computational and implementation complexity and short detection time, there also exist some challenges in designing a good energy detector. First, the detection threshold depends on the noise power, which may change over time and hence is difficult to measure precisely in real time. In low-SNR regimes where the noise power is very high, reliable identification of a primary user is not even possible [424]. Moreover, an energy detector can only decide the primary user's presence by comparing the received signal energy with a threshold; thus, it cannot differentiate the primary user from other unknown signal sources. Hence, it can trigger false alarms frequently.

1.3.2.2 Feature detectors

There are specific features associated with the information transmission of a primary user. For instance, the statistics of the transmitted signals in many communication paradigms are periodic because of the inherent periodicities such as the modulation rate, carrier frequency, etc. Such features are usually viewed as cyclostationary features, on the basis of which a detector can distinguish cyclostationary signals from stationary noise. In a more general sense, features can refer to any intrinsic characteristics associated with a primary user's transmission, as well as the cyclostationary features. For example, center frequencies and bandwidths [476] extracted from energy detection can also be used as reference features for classification and determining a primary user's presence. In this section, we will introduce cyclostationary feature detection followed by a generalized feature detection.

Cyclostationary feature detection was first introduced in [132]. Since, in most communication systems, the transmitted signals are modulated signals coupled with sine-wave carriers, pulse trains, hopping sequences, or cyclic prefixes, while the additive noise is generally wide-sense stationary (WSS) with no correlation, cyclostationary feature detectors can be utilized to differentiate noise from primary users' signals [47] [318] [56] [376] [71] [353] [319] [144] [188] [222] [392] [397] [416] and distinguish among different types of transmissions and primary systems [250]. Features of the primary user's signal can be extracted by a cyclostationary detector to aid real-time detection of the primary user. For instance, in a spectrum-pooling system where

secondary users rent the temporarily used licensed bands, secondary users should monitor the channel-allocation information (CAI) continuously in order to vacate the frequency bands upon the appearance of a primary user. Exploiting the different cyclostationary properties of the primary and secondary signals [319] provides secondary users with immediate awareness of a primary user.

In contrast to an energy detector, which uses time-domain signal energy as test statistics, a cyclostationary feature detector performs a transformation from the time domain into the frequency feature domain and then conducts a hypothesis test in the new domain. Specifically, define the cyclic autocorrelation function (CAF) of the received signal $y(t)$ by

$$R_y^{\alpha}(\tau) = E[y(t+\tau)y^*(t-\tau)\mathrm{e}^{\mathrm{j}2\pi\alpha t}], \tag{1.8}$$

where $E[\cdot]$ is the expectation operation, $*$ denotes complex conjugation, and α is the cyclic frequency. Since periodicity is a common property of wireless modulated signals, while noise is WSS, the CAF of the received signal also demonstrates periodicity when the primary signal is present. Thus, we can represent the CAF using its Fourier-series expansion, called the cyclic spectrum density (CSD) function, expressed as [132]

$$S(f,\alpha) = \sum_{\tau=-\infty}^{\infty} R_y^{\alpha}(\tau)\mathrm{e}^{-\mathrm{j}2\pi f\tau}. \tag{1.9}$$

The CSD function has peaks when the cyclic frequency α equals the fundamental frequencies of the transmitted signal $x(t)$, i.e., $\alpha = k/T_x$ with T_x being the period of $x(t)$. Under hypothesis $\mathcal{H}_0$, the CSD function does not have any peaks since the noise is non-cyclostationary signals. A peak detector [141] or a generalized likelihood ratio test [318] [250] can be further used to distinguish between the two hypotheses. Different primary communication systems using different air interfaces (modulation, multiplexing, coding, etc.) can also be differentiated by their different properties of cyclostationarity.

However, when OFDM becomes the air interface, as suggested by several wireless communication standards, identification of different systems may become problematic, since the features due to the nature of OFDM signaling are likely to be similar or even identical. To address this issue, particular features need to be introduced into OFDM-based communications. In [272], methods that assign different properties of cyclostationarity to different systems are considered. The OFDM signal is configured before transmission so that its CAF outputs peak at certain pre-chosen cycle frequencies, and the difference in these frequencies is used to distinguish among several systems operating within the same OFDM air interface. A similar approach is considered in [392].

Compared with energy detectors that are prone to high false-alarm probability due to noise uncertainty and cannot detect weak signals in noise, cyclostationary detectors are good alternatives because they can differentiate noise from primary users' signal and have better detection robustness in a low-SNR regime. However, the computational complexity and the significant amount of observation time required for adequate

detection performance prevent wide use of this approach. A spectrum-sensing method based on maximum cyclic autocorrelation selection has been proposed in [267], where the peak and non-peak values of the cyclic autocorrelation function are compared to determine whether the primary signal is present or not. This method does not require noise-variance estimation, and is robust against noise uncertainty and interference signals. Frequency-selective fading and uncertain noise impair the robustness of cyclostationary signal detection in low-SNR environments. Run-time noise calibration has been considered in [424] and [423], in order to improve detector robustness. The method exploits the in-band measurements at frequencies where a pilot is absent to calibrate the noise statistics at the pilot frequencies. By combining neural network for signal classification with cyclic spectral analysis, a more efficient and reliable classifier is developed in [120]. Since a large amount of processing is performed offline using neural networks, the online computation for signal classification is greatly reduced.

Generalized feature detection refers to detection and classification that extracts more feature information other than the cyclostationarity due to the modulated primary signals, such as the transmission technologies used by a primary user, the amount of energy and its distribution across different frequencies [432] [280], the channel bandwidth and its shape [348] [476], the power spectrum density (PSD) [356], the center frequency [476], the idle guard interval of OFDM [230], an FFT-type feature [256], etc. By matching the features extracted from the received signal to the a priori information about primary users' transmission characteristics, primary users can be identified. Using the packet-length information extracted from the packet header [150], a continuous-time semi-Markov traffic model is developed that not only captures the WLAN's behavior but also helps in designing the optimal control policies for dynamic spectrum access.

Location information of the primary signal is also an important feature that can be used to distinguish a primary user from other signal sources. In primary-user-emulation attack, a malicious secondary user transmits signals whose characteristics emulate those of the primary signals. A transmitter verification scheme is proposed in [74] to secure trustworthy spectrum sensing based on location verification of the primary user. Different spectrum environments are analyzed and characterized in [269], which derives the closed-form probability distributions for the availability of a given frequency channel (narrowband and wideband) and energy appearing in the channel. Constructing a probability distribution of the spectrum environment enables researchers to perform spectrum analysis without requiring a large database and make provable assertions in various types of radio environments.

1.3.2.3 Matched filtering and coherent detection

If secondary users know information about a primary user's signal a priori, then the optimal detection method is matched filtering [349], since a matched filter can correlate the already-known primary signal with the received signal to detect the presence of the primary user and thus maximize the SNR in the presence of additive stochastic noise. The merit of matched filtering is the short time it requires to achieve a certain detection performance such as a low probability of missed detection and false alarms

[375] [458], since a matched filter needs fewer received signal samples. However, the required number of signal samples also grows as the received SNR decreases, so there exists an SNR wall [424] for a matched filter. In addition, its implementation complexity and power consumption are too high [71], because the matched filter needs receivers for all types of signals and corresponding receiver algorithms have to be executed. Matched filtering requires perfect knowledge of the primary user's signal, such as the operating frequency, bandwidth, modulation type and order, pulse shape, packet format, etc. If wrong information is used for matched filtering, the detection performance will be degraded a lot.

On the other hand, most wireless communication systems exhibit certain patterns, such as pilot tones, preambles, midambles, and spreading codes, which are used to assist control, equalization, synchronization, and continuity, or for reference purposes. Even though perfect information about a primary user's signal might not be attainable, if a certain pattern is known from the received signals, coherent detection (or waveform-based sensing) can be used to decide whether a primary user is transmitting or not [404] [82] [274]. As an example, the procedure of coherent detection using a pilot pattern is explained as follows [404].

There are two hypotheses in coherent detection:

$$\begin{aligned} \mathcal{H}_0 &: y(t) = n(t), \\ \mathcal{H}_1 &: y(t) = \sqrt{\epsilon} x_{\mathrm{p}}(t) + \sqrt{1-\epsilon} x(t) + n(t), \end{aligned} \tag{1.10}$$

where $x_{\mathrm{p}}(t)$ is a known pilot tone, ϵ is the fraction of energy allocated to the pilot tone, $x(t)$ is the desired signal, which is assumed to be orthogonal to the pilot tone, and $n(t)$ is additive white noise. The test statistic of the coherent detection is defined as the projected received signal in the pilot direction, i.e.,

$$T = \frac{1}{N} \sum_{t=1}^{N} y(t) \hat{x}_{\mathrm{p}}(t), \tag{1.11}$$

where $\hat{x}_{\mathrm{p}}$ is a normalized unit vector in the direction of the pilot tone. As N increases, the test statistic T under hypothesis $\mathcal{H}_1$ is much greater than that under $\mathcal{H}_0$. By comparing T with a predetermined detection threshold, one can decide on the presence of a primary user.

Coherent detection can also be performed in the frequency domain [356]. One can express the binary hypothesis test using the PSD of the received signal $S_Y(\omega)$, and distinguish between $\mathcal{H}_0$ and $\mathcal{H}_1$ by exploiting the unique spectral signature exhibited in $S_X(\omega)$. For instance, the PSD of the received signals can be estimated from a periodogram, which uses an n-point received signal to obtain the squared magnitudes of the n-point discrete-time Fourier transform $S_Y^{(n)}(k)$, $k = 0, 1, \ldots, n-1$. If the n-point sampled PSD of the signal $S_X^{(n)}(k)$ is determined at the receiver by exploiting the a priori known spectral features, the presence of a TV signal can be detected [356] using the following test statistics:

$$T_n = \frac{1}{N} \sum_{k=0}^{n-1} S_Y^{(n)}(k) S_X^{(n)}(k). \tag{1.12}$$

It is also shown that frequency-domain coherent detection with a priori known features can detect TV signals reliably from AWGN at very low SNR.

Coherent detection is shown to be robust against noise uncertainty, and not limited by the SNR wall [404] as long as N is large enough. Moreover, coherent detection outperforms energy detection in the sensing convergence time [82] [413], because the sensing time of energy detection increases quadratically with the SNR reduction, while that of coherent detection increases only linearly [413]. However, information about the waveform patterns is a prerequisite for implementing coherent detection; the more precise information a coherent detector has, the better the sensing performance will be. Moreover, utilization of a wide range of spectrum requires a frequency synthesizer, which, however, generates a square-wave local oscillator signal containing many harmonics (termed harmonic images) in addition to the fundamental frequency and degrades the performance of the primary-user detection. The use of a frequency offset to decorrelate the desired signals from the harmonic images is proposed in [285] in order to reject the harmonic images and improve spectrum sensing.

1.3.2.4 Other techniques

Several other spectrum-sensing techniques have been proposed in the recent literature, and some of them are variations inspired by the above-mentioned sensing techniques.

Statistical-covariance-based sensing. Since the statistical covariance matrices of the received signal and noise are generally different, the difference is used in [503] [501] to differentiate the desired signal component from background noise. The eigenvalues of the covariance matrix of the received signal can also be used for primary detection [502]. The ratio of the maximum eigenvalue to the minimum eigenvalue is quantized on the basis of random-matrix theory [425], and the detection threshold can be found among them. By the simulation on detecting digital TV signals, these methods based on statistical covariances have been shown to be more robust against noise uncertainty while requiring no a priori information about the signal, the channel, and the noise power.

Filter-based sensing. Application of a specific class of filter banks is proposed in [113] for spectrum sensing in CR systems. When filter banks are used for multicarrier communications in CR networks, the spectrum sensing can be performed merely by measuring the signal power at the outputs of subcarrier channels, with virtually no computational cost. The multitaper method [160] can also be thought of as a filter-bank spectrum estimation with multiple filter banks.

Fast sensing. By utilizing the theory of quickest detection, with which one performs a statistical test to detect the change of distribution in spectrum-usage observations as quickly as possible, agile and robust spectrum sensing is achieved in [251]. The unknown parameters after a primary user appears can be estimated using the proposed successive refinement, which combines both generalized likelihood ratio and parallel cumulative-sum tests. An efficient sensing sequence is developed in [239] to reduce

the delay due to spectrum-opportunity discovery. The probability that a frequency band is available at sensing, the sensing duration, and the channel capacity are three factors that determine the sensing sequence.

Learning/reasoning-based sensing. An approach based on reinforcement learning for the detection of spectral resources in a multi-band CR scenario is investigated in [24], where the optimal detection strategy is obtained by solving a Markov decision process (MDP). A MAC layer spectrum-sensing algorithm using knowledge-based reasoning is proposed in [446], where the optimal range of channels to finely sense is determined through proactive fast sensing and channel-quality information.

Measurements-based sensing and modeling. By collecting data over a long period of time at many base stations, [455] provides a unique analysis of cellular primary usage. The collected data are dissected along different dimensions to characterize the primary usage, and it is found that a random-walk process can be used to model the aggregate cell capacity, while the commonly adopted exponential distribution is not a good model for call durations. With the aid of a spectrum observatory, [22] extends short-term spectrum-usage measurements to study the spectrum-usage trend over long periods, observes spectrum-usage patterns, and detects the positions of spectrum white space in the time and spatial domains. Such information can be greatly helpful in developing good dynamic-access protocols and governing secondary systems.

Hough-transform-based sensing. The Hough transform, which has been studied in the image-processing literature for detection of patterns in binary images, can be used for detecting patterns in primary-user signals such as radar pulses [72], as long as the radio signals exhibit periodic patterns.

1.3.3 Cooperative sensing

The performance of spectrum sensing is limited by noise uncertainty, shadowing, and multipath fading effects. When the received primary SNR is too low, there exists a SNR wall, below which reliable spectrum detection is impossible even with a very long sensing time. If secondary users cannot detect the primary transmitter, while the primary receiver is within the secondary users' transmission range, a hidden-primary-user problem will occur, and the primary user's transmission will be interfered with.

By taking advantage of the independent fading channels (i.e., spatial diversity) and multiuser diversity, cooperative spectrum sensing is proposed to improve the reliability of spectrum sensing, increase the detection probability to better protect a primary user, and reduce the false-alarm rate to utilize the idle spectrum more efficiently. In centralized cooperative spectrum sensing, a central controller, e.g., a secondary base station, collects local observations from multiple secondary users, decides the available spectrum channels using some decision-fusion rule, and informs the secondary users which channels to access. In distributed cooperative spectrum sensing, secondary users exchange their local detection results among themselves without requiring a backbone infrastructure, and hence with reduced cost. Relays can also be used in cooperative spectrum sensing, such as the cooperative sensing scheme proposed in [155], where the cognitive users operating in the same band help each other relay information using

an amplify-and-forward protocol. It is shown that the inherent network asymmetry can be exploited to increase the agility. An extension to multiuser networks is studied in [156], and a decentralized cooperation protocol is proposed to ensure an agility gain for a large network population.

There also exist several challenges with cooperative spectrum sensing. For instance, secondary users can be low-cost devices equipped with only a limited amount of power, so they cannot afford very complicated detection hardware and high computational complexity. In wideband cooperative sensing, multiple secondary users have to scan a wide range of spectrum channels and share their detection results. This results in a large amount of sensory-data exchange, high energy consumption, and an inefficient data throughput. If the spectrum environment is highly dynamic, the sensed information may even be stale due to user mobility, channel fading, etc.

1.3.3.1 User selection

Owing to secondary users' different locations and channel conditions, it is shown in [341] that cooperation of all secondary users in spectrum sensing is not optimal, and the optimum detection/false-alarm probability is achieved by limiting the cooperation to a group of users who have relatively high SNR of the received primary signal.

Since detecting a primary user costs battery power of secondary users, and shadow fading may be correlated for nearby secondary users, an optimal selection of secondary users for cooperative spectrum sensing is desirable. In [403], various algorithms based on different amounts of available information are proposed to select a proper set of sensors that experience uncorrelated shadow fading. A joint spatial–temporal sensing scheme for CR networks is proposed in [102], where secondary users collaboratively estimate the location and transmitting power of the primary transmitter to determine their maximum allowable transmission power, and use the location information to decide which users should participate in collaborative sensing in order to minimize correlation among the secondary users. Performance evaluation of cooperative spectrum sensing over realistic propagation environments, i.e., correlated log-normal shadowing both in the sensing and in the reporting channel, is investigated in [106]. This work also provides guidelines to select the optimal number of users in order to guarantee a certain detecting performance in a practical radio environment.

In a CR sensor network, individual sensor nodes may experience a heterogeneous false-alarm and detection probability due to their different locations, making it harder to determine the optimal number of cooperative nodes. Sensor clustering is proposed in [238], where the optimal cluster size is derived so as to place an upper bound on the variation of the average received signal strength in a cluster of sensor nodes. Moreover, the sensor density is optimized so that the average distance between neighboring nodes is lower-bounded and their measurements are nearly independent, without much correlation.

If a secondary user cannot distinguish between the transmissions of a primary user and another secondary user, he will lose the opportunity to use the spectrum. It is shown in [404] that the presence/absence of possible interference from other secondary users is the main reason for the uncertainty in primary-user detection, and coordinating

nearby secondary users can greatly reduce the noise uncertainty due to shadowing, fading, and multipath effects. A good degree of coordination should be chosen on the basis of the channel coherent times, bandwidths, and the complexity of the detectors.

1.3.3.2 Decision fusion

Various decision-fusion rules for cooperative spectrum sensing have been studied in the literature. A logical OR rule is used [148] for combining multiple users' decisions for spectrum sensing in fading environments. Cooperative spectrum sensing using a counting rule is studied in [211], where sensing errors are minimized by choosing the optimal settings for both matched filtering and energy detection. It is shown in [506] that a half-voting rule is the optimal decision-fusion rule in cooperative sensing based on energy detection. Light-weight cooperation based on hard decisions is proposed [297] for cooperative sensing to alleviate the sensitivity requirements on individual users. A linear-quadratic (LQ) strategy has been developed [430] to combat the detrimental effects of correlation between different secondary users.

A good way to optimally combine the received primary-signal samples in space and time is to maximize the SNR of local energy detectors. However, optimal combination requires information about the signal and channel. Blindly combined energy detection is proposed in [505], which, without requiring such information and noise-power estimation, performs much better than an energy detector and is more robust against noise uncertainty. Hard decision combined with the logical AND rule and soft decision using the likelihood ratio test are proposed in [435] for use in collaborative detection of TV transmissions. It is shown that soft decision combining for spectrum sensing yields more precise detection than hard decision combining. Soft decision combination for cooperative sensing based on energy detection is investigated in [287], and maximal ratio combination (MRC) is proved to be near optimal in low-SNR regions and to reduce the SNR wall. A softened hard combination scheme with two-bit overhead is further proposed, which achieves a good tradeoff between detection performance and complexity.

In general, cooperative sensing is coordinated over a separate control channel, so a good cooperation scheme should be able to use a small bandwidth and power for exchanging local detection results while maximizing the detection reliability. An efficient linear cooperation framework for spectrum sensing is proposed in [354], where the global decision is a linear combination of the local statistics collected from individual nodes using energy detection. Compared with the likelihood ratio test, the proposed method has lower computational complexity, closed-form expressions for the detection and false-alarm probabilities, and comparable detection performance.

The performance of cooperative spectrum sensing depends on the correctness of the local sensing data reported by the secondary users. If malicious users enter a legitimate secondary network and compromise the secondary users, false detection results will be reported to the fusion center, and this kind of attack is called a spectrum-sensing data falsification (SSDF) attack [75]. In order to guarantee a satisfying detection performance under SSDF attack, a weighted sequential probability ratio test (WSPRT) is proposed in [75], which incorporates a reputation-based mechanism into the

sequential probability ratio test. If a secondary user's local detection result is identical to the final result after decision fusion, his/her reports will carry more weight in future decision fusion. The proposed WSPRT approach is more robust against SSDF attack than are commonly adopted decision-fusion rules, such as AND, OR, and majority rules [83].

1.3.3.3 Efficient information sharing

In order to coordinate the cooperation in spectrum sensing, a lot of information exchange among secondary users is needed, such as their locations, estimation of the primary user's location and power, which users should be clustered into a group, and which users should perform cooperate sensing at a particular time epoch. Such a large amount of information exchange brings a lot of overhead to the secondary users, which necessitates efficient information sharing among the secondary users.

The guess protocol, an incremental gossiping approach, is proposed in [4] as a means to coordinate the dissemination of spectrum-sensing results. It is shown that the proposed approach can reduce overhead because the amount of information exchange among secondary users is limited, and this method accommodates network alternations such as node movement or node failures and exhibits fast information convergence. In order to reduce the bandwidth required by a large number of secondary users for reporting their sensing results, a censoring method with quantization is proposed in [411]. Only users with reliable information will send their local observations, i.e., a one-bit decision 0 or 1, to the common receiver. It is shown that the sensing overhead is greatly reduced at the cost of a little degradation in detection performance. A pipelined spectrum-sensing framework is proposed in [157], where spectrum sensing is conducted concurrently while secondary users are sending their detection reports. The proposed method alleviates sensing overhead by making use of the reporting time, provides more time for spectrum sensing, and thus improves the detection performance. A multi-threaded sequential probability ratio test is further proposed for data fusion in the pipelined framework. Random matrix theory (RMT) is applied in cooperative spectrum sensing in [51]. The proposed method uses multiple secondary receivers to infer the structure of the primary signal using RMT without requiring information about the noise statistics or its variance. It is shown that the proposed method can estimate the spectrum occupancy reliably only with a small amount of received primary-signal samples.

1.3.3.4 Interference diversity

Traditional cooperative spectrum sensing schemes usually utilize the shadowing/multipath diversity among multiple secondary users to enhance the detection reliability, while the potential presence of an unknown number of low-power and time-varying interference sources is actually the main reason for noise uncertainty [334]. Since individual users have different local observation of those low-powered interference sources, e.g., the arrival or departure of one interfering source merely causes a few nearby sensors to trigger false alarms whereas a primary user's activity can be detected by many more secondary users far apart, such interference diversity is utilized to

develop an event-based cooperative sensing scheme for detecting the primary user [334].

1.3.3.5 Distributed cooperative sensing

Cooperative spectrum sensing has been shown to be able to greatly improve the sensing performance in CR networks. However, if cognitive users belong to different service providers, they tend to contribute less in sensing in order to increase their own data throughput. A distributed cooperating spectrum sensing scheme based on evolutionary game theory is proposed in [451] to answer the question of "how to collaborate" in multiuser de-centralized CR networks. Using replicator dynamics, the evolutionary game modeling provides an excellent means to address the strategic uncertainty that a user may face by exploring various actions, adaptively learning during the strategic interactions, and approaching the best response strategy under changing conditions and environments. The behavior dynamics and the optimal cooperation strategy of the secondary users are characterized. A distributed learning algorithm is further developed so that the secondary users approach the optimal strategy solely on the basis of their own payoff observations. The proposed game is demonstrated to achieve a higher system throughput than the fully cooperative scenario, where all users contribute to sensing in every time slot. Another form of throughput-enhancing cooperative spectrum sensing is proposed in [264], where secondary users share their decisions about the spectrum occupancy of the primary users, and have more opportunities for access to idle spectrum with fewer collisions with primary users. The proposed scheme requires a common control channel, and can work in a distributed fashion.

1.3.3.6 Experimental measurements

Cooperative sensing using energy detection has been implemented on a wireless testbed [81]. Experimental study has demonstrated the improvement in detection performance due to cooperation, such as the need for less sensing time and the achievement of a higher detection probability. A measurement setup for cooperative spectrum sensing and experimental results obtained with it are presented in [459]. It is shown by measurements that the cooperation gain increases with the distance between secondary users and that cooperative sensing can alleviate the sensitivity requirements on a single secondary user. Moreover, since correlated spectrum measurements degrade the gain of cooperation, a robust metric for evaluating the correlations is developed.

1.4 Dynamic spectrum allocation and sharing

In the previous section, we have discussed various detection techniques and how to perform efficient cooperative spectrum sensing in order to obtain an accurate estimation of the interference temperature and spectrum-occupancy status. With the detection results, a secondary user will have an idea regarding which spectrum bands he/she could use. However, the availability and quality of a spectrum band may change rapidly with time due to primary users' activity and competition from other secondary users. In order to

Table 1.2. Classification of spectrum-allocation and -sharing schemes

Classification criterion	Type 1	Type 2
Spectrum bands that secondary users are using	Open spectrum sharing: access unlicensed spectrum band only	Hierarchical access/licensed spectrum sharing: also access licensed spectrum band
Access technology of licensed spectrum sharing	Spectrum underlay: secondary users transmit concurrently with primary users subject to interference constraints	Spectrum overlay: secondary users use the licensed spectrum only when primary users are not transmitting
Network architecture	Centralized: a central entity controls and coordinates the spectrum allocation and access	Distributed: each user makes his/her own decision on the spectrum-access strategy
Access behaviors	Cooperative: all secondary users work toward a common goal	Noncooperative: different users have different objectives

utilize the spectrum resources efficiently, secondary users need to be able to address issues such as when and how to use a spectrum band, how to coexist with primary users and other secondary users, and which spectrum band they should sense and access if the one currently in use is not available. Therefore, in this section, we will review the existing approaches to spectrum allocation and sharing that answer these questions.

Before going into the details, we would like to briefly discuss the classification of the current schemes for spectrum allocation and sharing. The existing schemes can be classified according to various criteria, as summarized in Table 1.2.

The first classification is according to the spectrum bands that secondary users are using. Spectrum sharing among the secondary users who access the unlicensed spectrum band is referred to as *open spectrum sharing*. One example is the open spectrum sharing in the unlicensed industrial, scientific, and medical (ISM) band. In open spectrum sharing, since no users own spectrum licenses, they all have the same rights in using the unlicensed spectrum. Spectrum sharing among the secondary users and primary users in licensed spectrum bands is referred to as the *hierarchical access model* [507] or licensed spectrum sharing. Primary users, who are usually not equipped with CRs, do not need to perform dynamic/opportunistic spectrum access, since they have priority in using the spectrum band. Whenever they reclaim the spectrum usage, secondary users have to adjust their operating parameters, such as power, frequency, and bandwidth, to avoid interrupting the primary users.

Considering the access technology of the secondary users, licensed spectrum sharing can be further divided into two categories [7] [507].

(i) *Spectrum underlay*. In spectrum underlay secondary users are allowed to transmit their data in the licensed spectrum band when primary users are also transmitting. The interference-temperature model is imposed on secondary users' transmission power so that the interference at a primary user's receiver is within the interference-temperature limit and primary users can deliver their packet to the

receiver successfully. Spread-spectrum techniques are usually adopted by secondary users to fully utilize the wide range of spectrum. However, due to the constraints on transmission power, secondary users can achieve only short-range communication. If primary users transmit data all the time in a constant mode, spectrum underlay does not require secondary users to perform spectrum detection to find available spectrum band.

(ii) *Spectrum overlay*. Spectrum overlay is also referred to as opportunistic spectrum access. Unlike spectrum underlay, secondary users in spectrum overlay will use the licensed spectrum only when primary users are not transmitting, so there is no interference-temperature limit imposed on secondary users' transmission. Instead, secondary users need to sense the licensed frequency band and detect the spectrum white space, in order to avoid harmful interference with primary users.

The second classification [7] is according to the network architecture. When there exists a central entity that controls and coordinates the spectrum allocation and access of secondary users, the spectrum allocation is *centralized*. If there is no such central controller, perhaps because of the high cost of constructing an infrastructure or the ad hoc nature of the network such as for emergency or military use, that kind of spectrum sharing belongs to the category of *distributed* spectrum sharing. In distributed spectrum sharing, each user makes his own decision about his spectrum-access strategy, mainly on the basis of local observation of the spectrum dynamics.

The third classification is according to the access behavior of secondary users [7]. If all secondary users work toward a common goal, for instance they belong to the same operator or service provider, they will coordinate their allocation and access in order to maximize their social welfare. This is called *cooperative* spectrum sharing. Most forms of centralized spectrum allocation can be considered cooperative. On the other hand, it is not always the case that all secondary users belong to the same service provider; e.g., it is not the case for those who access the open spectrum band. Different users have different objectives, and hence they aim only at maximizing their own benefit from using the spectrum resources. Since users are no longer cooperating to achieve the same objective, this kind of spectrum sharing is *noncooperative*, and secondary users are selfish in that they pursue their own benefit.

In order to give the reader more insight into how to design efficient spectrum allocation and sharing schemes, we next discuss several important issues in dynamic spectrum allocation and sharing.

1.4.1 Medium-access control in CR networks

Medium-access control refers to the policy that controls how a secondary user should access a licensed spectrum band. Various medium-access control protocols have been proposed in wireless networking such as carrier-sense multiple access (CSMA) and slotted ALOHA. Owing to the new features of CR networks, such as the requirement for spectrum sensing and access to avoid collision with a primary user, dynamics in

spectrum availability, and adaptation to a changing environment, new medium-access protocols need to be designed to address new challenges in CR networks.

A cognitive medium-access protocol with stochastic modeling is proposed in [152], which enhances the coexistence of CR with WLAN systems that are based on sensing and prediction. The continuous-time Markov chain is adopted to approximate the primary user's traffic, and a sense-before-transmit strategy constrains the interference generated toward the primary user. The CR's throughput is optimized by solving a constrained Markov decision process using linear programming. An implementation of the cognitive medium-access protocol is presented in [149]. A primary-prioritized Markov approach for dynamic spectrum access is proposed in [449], which models the interactions between the primary users and the secondary users as continuous-time Markov chains. By designing appropriate access probabilities for the secondary users, a good tradeoff between spectrum efficiency and fairness, and a higher throughput than with CSMA-based random access, can be achieved. A cognitive MAC (C-MAC) protocol for distributed multi-channel wireless networks is introduced in [48]. Since the C-MAC operates in multiple channels, it is able to deal with the dynamics of channel availability due to primary users' activity. Beaconing is included in a frame so that users can exchange local information, negotiate channel usage, and avoid hidden-user problems. A distributed and dynamic coordination among nodes in different channels can be achieved using a rendezvous channel. A stochastic channel-selection algorithm based on learning automata is proposed in [381]. This dynamically adapts the probability of access to one channel in real time and asymptotically converges to the optimal channel. It is shown that the probability of successful transmissions is maximized using the proposed selection algorithm.

Opportunistic scheduling policies for CR networks using the technique of Lyapunov optimization are investigated in [429]. This technique maximizes the throughput of secondary users while upper-bounding the collisions with primary users. A Multi-MAC protocol that can dynamically reconfigure MAC- and physical-layer properties for CR networks is proposed in [105]. This protocol, which is based on per-node and per-flow statistics, provides intelligent reconfiguration of the MAC and physical layers in response to environment variations, and hence achieves the best performance while ensuring correct decoding of incoming frames using the proper MAC-layer algorithm. The impact of channel heterogeneity, such as the heterogeneity in transmission ranges, data rates, etc., on network performance is identified in [243], which motivates the need to account for channel heterogeneity in designing higher-layer protocols. Considering the limited capability of spectrum sensing and limited bandwidth, a hardware-constrained cognitive MAC is proposed in [218], which optimizes the spectrum-sensing decision by formulating sensing as an optimal-stopping problem. Using backward induction, the optimal sensing strategy is derived on the basis of observations of reward from the past.

Secondary users also need to be aware of their surrounding environment in allocating and accessing the spectrum. Considering that each node's spectrum usage is unpredictable and unstable, the work in [89] proposes integrating interference-aware statistical admission control with stability-oriented spectrum allocation. The

nodes' spectrum demand is regulated to allow efficient statistical multiplexing while the outage is minimized. Near-optimal algorithms are developed to solve the NP-hard spectrum-allocation problem. An efficient opportunistic access scheme should achieve a high data rate of secondary users while sufficiently protecting the primary user from harmful interference. Since secondary users operating in different frequency bands at different locations are constrained by different interference requirements, a good spectrum access scheme needs to take the interference heterogeneity into consideration. A distance-dependent MAC protocol is proposed in [389] to optimize the CR network throughput subject to a power-mask constraint to protect the primary user. The protocol adopts a probabilistic channel assignment algorithm, which exploits the dependence of the signal's attenuation model on the transmission distance while considering the local traffic profile. An idea of how to utilize location awareness to facilitate spectrum sharing between secondary and primary users is illustrated in [439]. With the development of discontiguous orthogonal frequency-division multiplexing, discontiguous spectrum access and spectrum aggregation has become possible, since spectrum fragments could be aggregated and further utilized. An aggregation-aware spectrum-assignment scheme is proposed in [90] to optimize the spectrum assignment when the available spectrum band is not contiguous.

The modeling of opportunistic spectrum access considering the dynamic traffic pattern is usually based on queuing theory, since queuing theory provides a systematic analytic tool for studying the performance in terms of packet delay, buffer length, system throughput, and so on. A queuing analytic framework is developed in [362] to evaluate the performance of secondary users in a CR network, such as queuing delay and buffer statistics of secondary users' packets. A channel-allocation scheme based on the queuing analytic model is considered. This approach can guarantee a required statistical delay performance. The collision probability and overlapping time are introduced in [176] to evaluate the protection of a primary user. With constraints on sufficient primary-user protection, various spectrum access schemes using different sensing, back-off, and transmission mechanisms are presented, which reveal the impact of several important design criteria, such as sensing, packet-length distribution, back-off time, packet overhead, and grouping.

1.4.2 Spectrum handoff

When the current channel conditions become worse, or the primary user appears and reclaims his assigned channel, secondary users need to stop transmitting data and find other available channels in which to resume their transmission. This kind of handoff in CR networks is termed *spectrum handoff* [7]. Since the transmissions of secondary users are suspended during spectrum handoff, they will experience longer packet delay. Therefore, a good spectrum handoff mechanism should provide secondary users with a smooth frequency shift with the least possible latency.

A good way to alleviate the performance degradation due to long delay is to reserve a certain number of channels for potential spectrum handoff [508]. When secondary users need to switch to another frequency, they can immediately pick one channel from

the reserved bands. However, if a secondary user reserves too much bandwidth for spectrum handoff, the throughput may be unnecessarily low, because the primary user might not reclaim his licensed band very frequently. Therefore, there is a tradeoff in optimizing the channel reservation. Assuming the arrival and service processes are Poisson, the process of spectrum occupation is modeled as a continuous-time Markov chain [508], with which the probability of service disruption due to primary users' appearance can be calculated. By optimizing the number of channels reserved for spectrum handoff, the blocking probability can be minimized and the secondary users' throughput is maximized. A location-assisted handover algorithm is proposed in [44]. A set of candidate channels is maintained by a secondary base station. Secondary users equipped with location-estimation and sensing devices can report their locations back to the secondary base station. Whenever handoff becomes a must, secondary users can switch their frequency to one of the candidate channels, depending on their locations. The algorithm can reduce the packet error rate due to primary users' activity and channel impairments, alleviate signaling overhead in choosing new available channels in real time, and maintain effective handoff and seamless communication. In a multi-hop CR network, the question of how to design a spectrum handoff mechanism becomes more complicated, because multiple links are involved. A joint spectrum handoff scheduling and routing (JSHR) protocol in multi-hop multi-radio CR networks is proposed in [117], which extends the spectrum handoff of a single link to that of multiple links. The JSHR problem is formulated so as to minimize the total handoff latency under the constraint on network connectivity. A distributed greedy algorithm is developed to solve the NP-hard problem, and a rerouting mechanism with spectrum-handoff scheduling is further designed to improve the network throughput. Three types of spectrum handoff for link maintenance have been studied in [438], including evaluation of the link-maintenance probability and effective throughput. Numerical results show that the probability of erroneous channel selection, the radio sensing time, and the number of handoff trials are important for spectrum-handoff design.

In order to achieve reliable continuous communication among secondary users in the presence of random reclaims from a primary user, secondary users should select their channels from different licensed bands owned by different primary users [444]. Such multi-band spectrum diversity helps to reduce the impact of the appearance of a primary user and improve the reliability of secondary spectrum access. Moreover, through adding some redundancy to the payload data, secondary users' transmission will be made robust against errors due to primary users' corruption. Multi-band diversity has also been suggested in [244], where the multimedia content is distributed over multiple temporarily idle spectrum bands. However, the nature of multimedia content distribution requires packet-scheduling strategies that ensure the quality of the received data; this will complicate the system design in a secondary-user environment where the arrival of a primary user is unpredictable. Therefore, the authors of [244] propose the usage of digital fountain codes, which not only helps multimedia content distribution to secondary users without coordination among them, but also combats the packet loss due to a primary user's activity and channel impairments. Optimal channel selection to meet the QoS requirement of multimedia streaming is discussed too. Luby transform

(LT) codes are proposed in [228] to compensate for the loss caused by the primary-user interference, and the optimal number of channels that maximizes the secondary users' spectral efficiency, given fixed parameters of the LT code, is studied.

It is also suggested that secondary users transmit concurrently with the primary users in the licensed band while mitigating interference to the primary users by coding techniques. A joint coding and scheduling method for cognitive multiple access is proposed in [68]. A successive interference decoder is utilized in the physical layer to mitigate the secondary user's interference with the primary user, and thus the secondary user is allowed to share a channel with the primary user. A joint channel-aware and queue-aware scheduling protocol is proposed in the MAC layer to minimize the secondary user's delay with given power constraints. With the proposed approach, secondary users do not have to wait until the end of the primary users' duty cycle to start transmission, and thus this approach improves the spectrum utilization with reduced delay.

1.4.3 Cognitive relaying

Cooperative relaying utilizing the broadcasting nature of wireless networks has been proposed in recent years [258] [257] as a means by which to improve the network performance through spatial and multiuser diversity. In cooperative relaying, relay nodes can forward a source node's data to a destination. Combined with CR technology, cooperative relaying can offer a more significant performance gain, because cognitive relay nodes can forward a source node's data by using the spectrum white space they have detected.

In wireless networks, a source node's traffic is in general bursty, so a cognitive relay can utilize the periods of silence of the source to enable cooperation. Motivated by this fact, the authors of [391] proposed a novel cognitive multiple-access strategy in the presence of a cooperating relay. Since the cognitive relay forwards data only when the source is not transmitting, no extra channel resources are allocated for cooperation at the relay, and hence the proposed protocols provide significant performance gains over conventional relaying strategies. By exploiting source burstiness, secondary users utilize primary users' periods of silence to access the licensed spectrum and help primary users forward their data, which not only increases their channel access probabilities but also achieves a higher throughput [112]. A cognitive OFDM-based spectrum-pooling system is considered in [337]. The source node transmits data to the relay using a certain pool of subcarrier frequencies, and the relay forwards the received data to the destination node on a possibly different pool of OFDM subcarriers. By determining an optimum assignment of the subcarrier pools, the capacity of the spectrum-pooling relay system is maximized. A frequency-sharing multi-hop CR network is studied in [127]. By recognizing the radio environment in each relay node, the system can autonomously avoid transmission in an interference area. In [219], an infrastructure-based secondary network architecture is proposed to leverage relay-assisted discontiguous OFDM for data transmission. A relay node that can bridge the source and the destination using its common channels between the two nodes will be selected. Relay selection and spectrum

allocation are jointly optimized, and a corresponding MAC protocol is proposed and implemented in a Universal Software Radio Peripheral-based testbed.

There are also several other works that study the performance of cognitive relay networks, including the achievable region and outage probability. The achievable region for a two-source, two-destination Gaussian interference channel with a cognitive relay is studied in [407], where the cognitive relay has access to messages transmitted by both sources and assists them in relaying the messages to their respective destinations. A one-sided interference channel assisted by a cognitive relay is considered in [380], where the relay has a link only to the destination that observes interference. Good relay strategies are investigated. It is found from the achievable region that, under certain conditions, the relay uses most of its power in canceling out the interference instead of boosting the desired signal power. The outage performance of cognitive wireless relay networks is studied in [263]. A group of network clusters consisting of several unlicensed relay nodes helps the source node forward messages to the destination node by using the spectrum white space. High-SNR approximation of the outage probability of the two-hop system is investigated, and it is found that full diversity is achieved only if each relay node successfully detects the spectrum hole. An intra-cluster cooperation scheme to improve the outage performance, in which appropriately many neighboring cognitive relay nodes inside a cluster collaborate with a desired cognitive relay node, is proposed too.

1.4.4 Spectrum sensing and access

Owing to energy and hardware constraints, a secondary user might not be able to sense the entire spectrum space and can access only a limited number of channels from those it has sensed. To optimize spectrum access while considering physical-layer spectrum sensing and the primary user's traffic statistics, a decision-theoretic approach based on a partially observable Markov decision process (POMDP) is proposed in [509]. The proposed method is shown to be able to optimize secondary users' performance, accommodate spectrum-sensing error, and protect primary users from harmful interference. A joint design and separation principle for opportunistic spectrum access using POMDP is proposed in [94]. The separation principle reveals the optimality of myopic policies for the spectrum-sensor design and access strategy, and reduces the complexity of the POMDP formulation by decoupling the design of the sensing strategy from the design of the access strategy. By exploiting the mixing time of the underlying Markov process of spectrum occupancy, a truncated MDP formulation is developed in [107], which provides a tradeoff between performance and computational complexity. DSA with perfect and imperfect sensing based on Markov chain modeling is studied in [445]. Aspects of system performance such as airtime and blocking probabilities are evaluated, and the impact of the false-alarm and misdetection probabilities on DSA is analyzed.

An extension of [509] that incorporates the secondary user's residual energy and buffer state into the POMDP formulation for spectrum sensing and access is presented in[93] [95]. Monotonicity results are developed: first, the secondary user with data to transmit should sense a channel if and only if the conditional probability of

the channel being idle is above a certain threshold; second, the secondary user should transmit over an idle channel if and only if the channel fading is below a certain threshold. This observation can accelerate the decision making of a secondary user about the strategy for spectrum sensing and access.

Owing to energy and hardware constraints, secondary users need to choose carefully which bands they should sense and access. Continuously accessing the channel with the highest estimated chance of availability may bring short-term gain, whereas exploration enables the secondary users to learn the statistical behavior of the primary traffic, with long-term gain. Therefore, cognitive medium access is modeled as a multi-armed-bandit problem in [247], and an efficient access strategy is developed that achieves a good balance between exploring the availability of other free bands and exploiting the opportunities that have been identified. A multi-cognitive-user scenario is also considered, which is modeled as a game. A similar idea has been proposed in [499]. By formulating the process of spectrum sensing and access as a multi-arm-restless-bandit process, the authors of [499] studied the structure, optimality, and performance of the myopic sensing policy. It is shown that the myopic policy reduces channel selection to a round-robin procedure and alleviates the requirement on knowing the channel's state-transition probability. The maximum throughput of a multi-channel opportunistic system using the myopic sensing policy and its scaling behavior with respect to the number of channels are also characterized.

1.4.5 Power control in CR networks

Power control is a common approach for alleviation of interference. In order to manage the interference among secondary users, or avoid harmful interference with primary users due to secondary spectrum usage, various power-control schemes to coordinate spectrum sharing are also considered in CR networks.

Power control in opportunistic spectrum access (OSA) has been studied in [373], which models the packet transmission from source to destination in OSA as crossing a multi-lane highway. If a secondary user tries to use high transmission power to reach the destination in one hop, it has to wait until the primary user is inactive; on the other hand, it can take more advantage of the spectrum opportunities with lower transmission while relying on the intermediate users on the path to destination. The impact of transmission power on the occurrence of spectrum opportunities is investigated in [373], and it is shown that the optimal transmission power of secondary users decreases monotonically with the traffic load of the primary network. Dynamic programming has been used in designing an optimal power- and rate-control strategy, in order to maximize the long-term average rate for a secondary user [153], with the constraints on the total energy budget of the secondary user and the interference from the secondary user with the primary user. The power-adaptation strategies that maximize the secondary user's SNR and capacity under various constraints are studied in [387]. An opportunistic power-control strategy that enables the cognitive user to maximize its transmission rate while guaranteeing that the outage probability of the primary user is not degraded is proposed in [85]. In the proposed method, the cognitive user transmits with its maximum

power when it senses that the primary channel is already in outage. When the primary channel is not in outage, it transmits with a fraction of its maximum power that ensures successful transmission of the primary user. A collaborative spectrum-sensing scheme that considers signal strength, localization, and collaboration in the presence of multiple co-channel primary and secondary transmitters is proposed in [309]. The allowed maximum transmitter power of a secondary user in a given channel is determined using a distributed database containing co-channel transmitter information including location, error estimates, power, etc.

In most power-control schemes for CR networks, there usually exist interference constraints that prohibit simultaneous transmission by users that are within each other's transmission range. The interference constraints should be characterized clearly in order to reflect the interference relationship; in addition, the description should not be too complicated, otherwise a closed-form solution cannot be obtained easily. Therefore, a conflict graph is commonly adopted to describe the interference constraints among users, such that a node in the graph represents a user, and an edge between a pair of nodes represents the existence of interference. A multi-channel contention graph is proposed [419] to characterize the interference in a protocol-interference model, on the basis of which the spectrum allocation and scheduling in CR networks can be jointly optimized.

Most current works on DSA with interference constraints usually adopt the protocol model that simplifies interference constraints by presenting them as conflict graphs, which may suffer performance degradation due to incorrect interference estimation. A systematic framework to produce conflict graphs on the basis of a physical interference model is presented in [481], which characterizes the cumulative effect of interference while making it possible to use graph theory to solve spectrum-allocation problems under physical interference constraints.

1.4.6 Control-channel management

Most DSA systems use a dedicated global control channel to coordinate the spectrum allocation. However, this assumption is not realistic in opportunistic spectrum access since there might be no permanent channel available for secondary users. A distributed group-coordination solution is proposed in [513], where a common control channel is required only locally by the neighboring nodes sharing common channels. The concept of a segment is introduced in [35], where a group of nodes sharing common channels along a routing path coordinate the control channel selection. A cluster-based approach is presented in [92], where a dynamic one-hop cluster is formed by users sharing common channels and the spectrum is managed by cluster heads. A distributed swarm-intelligence-based control channel assignment scheme is proposed in [91], which selects local common control channels among a local group of secondary users according to the quality of the detected spectrum holes and the choice of the neighboring users.

In CR networks, control signals for coordinating spectrum sharing are transmitted through a dedicated channel, namely the common control channel (CCC). However,

potential control channel saturation will degrade the network performance severely. An alternative MAC protocol without requiring a CCC for multi-hop CR networks is proposed in [221]. By dividing the time into fixed time intervals and having all users listen to a channel at the beginning of each slot, the proposed protocol ensures that control signals can be exchanged among users. Simulation results show that the protocol provides higher throughput than that for a CCC-based protocol.

1.4.7 Distributed spectrum sharing

In centralized spectrum allocation, a lot of information needs to be exchanged among the central controller and network users to coordinate their spectrum usage, and this results in a large amount of signaling overhead. Therefore, distributed spectrum sharing is preferred where users can make their decisions on how to use the spectrum solely on the basis of local information.

A distributed spectrum management scheme is proposed in [88], where nodes take independent actions and share spectrum resources fairly. Five spectrum rules are presented to regulate node behavior. These rules are shown to achieve similar performance to that obtained with the explicit coordination approach while reducing the overhead due to information exchange. An adaptive approach to manage spectrum usage in dynamic spectrum-access networks is investigated in [87]. This approach achieves a comparable performance in spectrum assignment to that of the conventional centralized approach, with less information exchange. Considering the frequency agility and adaptive bandwidth, the concept of a time–spectrum block is introduced in [479], with which the spectrum-allocation problem is defined as the packing of time–spectrum blocks in a two-dimensional space. A distributed protocol is developed to solve the spectrum-allocation problem, which enables each node to dynamically choose the best time–spectrum block solely on the basis of local information. A biologically inspired spectrum-sharing algorithm based on the adaptive task-allocation model in insect colonies is introduced in [1]. The proposed algorithm enables secondary users to distributively determine the appropriate channels to use with no spectrum-handoff latency due to coordination, and achieves efficient spectrum sharing. A distributed resource-management algorithm that allows network nodes to exchange information and learn the actions of interfering nodes using a multi-agent learning approach is proposed in [406].

1.4.8 Spectrum sharing games

Game theory is a well-developed mathematical tool that studies the intelligent behaviors of rational decision makers in strategic interactions, such as cooperation and competition. In dynamic spectrum sharing, secondary users compete for the limited spectrum resources. If they do not belong to the same network entity, secondary users aim only at maximizing their own benefit from utilizing the spectrum resources. Therefore, their strategies in dynamic spectrum sharing can be well analyzed via game-theoretic approaches [206].

A game-theoretic modeling that analyzes the behavior of cognitive users in distributed adaptive channel allocation is presented in [304]. Both cooperative and non-cooperative scenarios are considered, and a no-regret learning approach is proposed. It is shown that cooperation-based spectrum sharing etiquette improves the overall performance at the cost of a higher overhead due to information exchange. In [110], a repeated-game approach for spectrum allocations is proposed, in which the spectrum sharing strategy could be enforced using the Nash equilibrium of dynamic games. A mechanism design to suppress the cheating behavior of secondary users in open spectrum sharing by introducing a transfer function into the user's utility is proposed in [466] [462]. The transfer function represents the payment that a user receives (or makes if it is negative) on the basis of the private information he/she announces in the spectrum-sharing game. In the proposed mechanism, it is shown that users can attain the highest utility only by announcing their true private information. A random-access protocol based on continuous-time Markov models for dynamic spectrum access in open spectrum wireless networks, is investigated in [468], where the secondary users' traffic arrival/departure process is assumed to be a Poisson random process. A distributed implementation of the protocol which controls the secondary users' access probability on the basis of a homo equalis model is proposed to achieve airtime fairness among them. Spectrum sharing among one primary user and multiple secondary users is formulated as an oligopoly market competition [305], and a Cournot game approach is proposed to obtain the spectrum allocation for secondary users, in which each secondary user's strategy is chosen on the basis of pricing information obtained from the primary user. The spectrum pricing problem when multiple primary users compete with each other to sell spectrum bands to secondary users is studied in [306], and a distributed algorithm to obtain the solution to the problem is presented. The game can achieve the highest total profit under a punishment mechanism that deters the primary users from deviating from the optimal solution. In dynamic secondary access, the accumulative amount of power from the secondary users should not violate the interference-temperature limit. With this constraint, a dynamic spectrum access optimization problem is formulated in [472] that can also guarantee a certain secondary QoS. A secondary spectrum-sharing potential game model is further proposed to solve the problem in a distributed fashion, using distributed sequential play and stochastic learning. A correlated equilibrium concept that can achieve better spectrum sharing performance than non-cooperative Nash equilibrium in terms of spectrum utilization efficiency and fairness is used in [186]. A no-regret learning algorithm is adopted to achieve the correlated equilibrium with proven convergence. A game-theoretic overview for dynamic spectrum sharing is provided in [206].

Auction mechanisms for spectrum sharing have also been proposed in [163]. Since users access the channel using spread spectrum signaling, they interfere with each other and have to allocate power carefully in order to utilize the spectrum more efficiently. Spectrum sharing among users is modeled as an auction, where the utility of each user is defined as a function of the received SINR. Considering the potential price of anarchy due to the non-cooperative nature of selfish users, the spectrum

manager charges each user a unit price for their received SINR or power. With the pricing introduced, the auction mechanism achieves the maximum social utility as well as maximal individual utility. An iterative bid-updating algorithm is also presented for the distributed implementation. A spectrum auction should consider carefully the local spectrum demand and spectrum availability in order to achieve high utilization. A real-time spectrum-auction framework is proposed in [134] to assign spectrum packages to proper wireless users under interference constraints. Different pricing models are considered in order to assess tradeoffs of revenue and fairness, and fast auction clearing algorithms are proposed to compute the revenue-maximizing prices and allocation. In [204] [207], a belief-assisted distributive pricing algorithm is proposed to achieve efficient dynamic spectrum allocation based on double-auction mechanisms, with collusion-resistant strategies that combat possible collusive behavior of users by using optimal reserve prices. A scalable multi-winner spectrum-auction scheme that awards one spectrum band to multiple secondary users with negligible mutual interference is proposed in [465]. Effective mechanisms to suppress dishonest/collusive behaviors are also considered, in case secondary users distort their valuations of spectrum resources and interference relationships. A truthful and computationally efficient spectrum auction is proposed in [495], which can support an eBay-like dynamic spectrum market and maintain truthfulness while maximizing spectrum utilization. A truthful double-auction mechanism is proposed in [512] to further increase spectrum efficiency by allowing spectrum reuse, since wireless users that do not interfere with each other can share the same spectrum bands.

1.4.9 Routing in CR networks

In traditional wireless networks, all network nodes will be provided with a certain fixed spectrum band for use. For instance, WLAN uses 2.4- and 5-GHz bands, and GSM uses 900- and 1800-MHz bands. In dynamic spectrum access (DSA) networks, however, there may be no such pre-allocated spectrum that can be used by every node at any time, and the frequency spectrum that can be used for communication may vary from node to node. This new feature of DSA networks imposes even greater challenges on wireless networking, especially on routing. If two neighboring nodes do not have a common channel, or they have common channels but do not tune to the same frequency, then multi-hop communication will not be feasible. Thus, new routing algorithms are needed in order to accommodate the spectrum dynamics and ensure satisfying network performance such as high network capacity and throughput, short latency, and low packet loss.

Owing to the heterogeneity of spectrum availability among nodes, the routing problem can not be well solved without considering the spectrum allocation. The interdependence between route selection and spectrum management is studied in [467], where two design methodologies are compared. The first is a decoupled approach in which route selection and spectrum management are performed independently in different protocol layers. The second approach is a collaborative design, in which some tasks of spectrum management are integrated into route selection in the network layer.

The network layer will select the packet route as well as decide a time schedule of a conflict-free channel usage. Experimental results show that a well-provisioned collaborative design outperforms the decoupled design.

In [474], the topology formation and routing in DSA networks is studied. DSA network nodes first identify spectrum opportunities by detection, and then the detected spectrum opportunities are associated with the radio interfaces of each node. A layered graph model to help assign the spectrum opportunities to the radio interfaces is proposed. Using the model, a routing path between nodes can be computed conveniently for each pair of nodes, which not only diversifies channel selection to prevent interference between adjacent hops along the path but also maximizes network connectivity.

A MAC-layer configuration algorithm that enables nodes to dynamically discover the global network topology and node location, and identify common channels for communication, is proposed in [241]. When a CR network can utilize multiple channels for parallel transmission, while the available channels vary with primary users' activity, traditional routing metrics such as energy consumption, number of hops, congestion, etc., are not sufficient for correct routing decision making. New routing metrics are introduced in [241], such as the number of channel switches along a path, frequency of channel switches on a link, and switching delay. Routing strategies to find the best route according to these new metrics in a CR network are proposed. Other routing metrics that incorporate the primary usage pattern, CR link hold-time, and throughput loss of primary users due to interference are considered in [260]. The capacity (per unit of time) of the links, the available spectrum, the link-disruption probabilities, and the link propagation time between nodes are considered for choosing a proper route in [327].

A spectrum-aware on-demand routing protocol is proposed in [69] [70]. This protocol selects routes according to the switching delay between channels and the back-off delay within a channel. A local coordination scheme is further proposed in [480], in which the intersecting nodes perform data-flow redirection according to the cost evaluation of frequency band switching and queuing delay. A probabilistic path-selection approach is proposed for multi-channel CR networks in [224]. The source node first computes the route that has the highest probability of satisfying a required demand, and then verifies whether the capacity of the potential path does indeed meet the demand. If not, extra channels are judiciously added to the links of the route until the augmented route satisfies the demand at a specified confidence level.

The returns of a primary user to a licensed band can also be viewed as a constraint on channel switching, since some channels will become locally unusable when the primary user appears. Considering the channel-switching constraints due to primary users' activity, the authors of [42] propose analytic models for channel assignment in a general multi-hop CR network, studies the impact of the constraints on network performance, and investigates the connectivity and transport capacity of the network. An analytic model and optimization framework for spectrum sharing in CR networks is proposed in [191], which considers the constraints on routing, flow control, interference, and capacity.

A cognitive networking architecture and a preliminary prototype and experimental setup of a cognitive-network access point are presented in [295]. The access point can obtain spatial and temporal patterns of higher-layer network traffic by real-time monitoring, which can be used for routing in CR networks.

A wireless mesh network (WMN) can be seen as a special type of wireless ad hoc network. A WMN usually consists of mesh clients (MCs), mesh routers (MRs), and gateways, which are organized in a mesh topology. The MCs are often laptops, mobile users, or other wireless devices, which direct their traffic to the respective MRs. The MRs, which form the backbone of the network, can be viewed as access points that forward the MCs' traffic over the backbone to and from the gateway in a multi-hop fashion. When one mesh node can not function well and communicate, the remaining nodes can still communicate with each other. Therefore, WMNs provide users with reliable communication and fault tolerance, as well as flexible network architectures and easy deployment. However, the network capacity will be reduced significantly when the node density per transmission channel increases and the network becomes congested. Hence, there is a strong need for rich spectrum resources to support the operation of WMNs, and opportunistic spectrum access (OSA) with CR has become an attractive solution [45]. Equipped with CR, the MCs can monitor the primary channels and identify the spectrum white space. The interference due to the mesh traffic at any frequency in any location can be estimated. An integer linear program is further formulated to solve the channel-assignment problem so that the MCs can fully utilize the idle licensed spectrum under certain interference constraints. The distributed approach for channel selection is scalable and also satisfies the interference requirement from primary users.

Moreover, the appearance of spectrum holes is highly dependent on location and time, and the available spectrum in each mesh node may be different. Two neighboring nodes cannot communicate with each other if they do not have a common channel or they are not tuning into the same channel. Therefore, mesh nodes should have knowledge of the available spectrum frequencies, scheduling, and routing path so that they can communicate with a minimal cost and no collision. An optimal two-hop spectrum scheduling in cognitive WMN is proposed in [475], where any pair within a two-hop neighborhood knows the spectrum allocation, collision-free scheduling, and minimal-cost routing path.

QoS routing in a cognitive WMN with interference constraints and dynamic channel availability is studied in [193]. This also is based on an integer linear programming formulation. A distributed routing protocol is developed that can optimally select a route and allocate channels and time slots to satisfy the end-to-end bandwidth requirement.

Owing to the heterogeneity of primary users' random behavior, if some node is severely affected by primary users' activity, that node should not be selected on a routing path. The approach proposed in [410] formulates the stochastic traffic engineering problem to address the issue of how the mesh traffic in the multi-hop cognitive WMN should be routed. Channel assignment with route discovery in cognitive WMN is also discussed in [130].

1.4.10 Security in CR networks

Owing to their new characteristics, such as the requirement on the awareness of the surrounding environment and internal state, reasoning and learning from observations and previous experience to recognize environment variations, adaptation to the environment, and coordination with other users/devices for better operation, CR networks face unique security challenges. In [40], awareness spoofing and its impact on different phases of a cognitive cycle have been studied. Through spoofing, the malicious attackers can cause an erroneously perceived environment, introduce biases to CR decision-making process, and manipulate secondary users' adaptation. In [76], the authors have investigated the primary-user emulation attack, where the cognitive attackers mimic the primary signal to prevent secondary users from accessing the licensed spectrum. A localization-based defense mechanism is proposed. This verifies the source of the detected signals by observing the signal characteristics and estimating its location from the received signal energy. The authors of [75] investigated the spectrum-sensing data-falsification attack, and proposed a weighted sequential probability ratio test to alleviate the performance degradation due to sensing error. In the proposed approach, individual sensing reports are compared with the final decision. Users whose reports are identical to the final decision will have high reputation values, and their reports will then carry more weight in future decision fusion. Several types of denial-of-service attacks in CR networks have been discussed in [38], such as spectrum-occupancy failures when secondary users are induced to interfere with primary users, policy failures that affect spectrum coordination, location failures, sensor failures, transmitter/receiver failures, compromised cooperative CR, and common control channel attacks. Simple countermeasures are also discussed. How to secure a CR network by understanding identity, earning and using trust for individual devices, and extending the usage of trust to networking has been discussed in [55].

1.5 Cognitive radio platforms

Although a lot of approaches have been proposed to improve the performance of spectrum sensing and dynamic spectrum access and sharing, most of them merely focus on the theoretical modeling and analysis and few of them have been verified in a practical system. Take spectrum sensing as an example. The primary users, who are usually not equipped with CR functionality, are concerned that the secondary users will interfere with their operation harmfully. This could happen if the secondary users cannot reliably detect a primary user and start transmission, while the primary user is active in the licensed band. Even if the secondary user has detected the primary user, it may fail to switch its frequency to some other available spectrum band fast enough and thus create harmful interference with the primary user's transmission. Therefore, CR platforms need to be developed as real-world testbeds that can verify the theoretical analysis. In this section, we will first review the existing testbeds/platforms developed by some research institutes and industry, followed by a brief discussion about standardization of CR techniques.

1.5.1 Berkeley Wireless Research Center

The feasibility of CR usage to efficiently utilize the spectrum resources without causing interference with the primary user cannot be justified unless it is shown in a real working system or testbed that the interference due to secondary users' activity is sufficiently low. Researchers at the University of California, Berkeley have proposed an experimental setup based on the Berkeley Emulation Engine 2 (BEE2) platform [277] to compare different sensing techniques and develop metrics and test cases so as to measure the sensing performance. Specifically, a good CR system should provide sufficient protection to the primary user, in the sense that the CR can detect the primary user within a very short time, reliably detect the primary user with a high detection probability and a low false-alarm probability, and vacate the spectrum quickly after a correct detection. These metrics impose certain requirements on a CR testbed, including the capability to support multiple radios, the ability to connect various different front-ends to support different frequency ranges, the capability for physical/link-layer adaptation and fast information exchange for sensing and cooperation, and the capability to perform rapid prototyping. The BEE2 can meet these requirements and support the features for a CR testbed. The BEE2 board can connect up to 18 front-ends, which enables the experiments with multiple primary users. It can also be used to perform complex signal processing with the aid of FPGAs, and the high-speed links between the FPGAs foster cooperation emulation among the secondary users.

Using the BEE2 platform, research on spectrum sensing using energy detection and sensing with cooperation was tested by experiments in [81], which shows the feasibility and practical performance limits of energy detection under real noise and interference in wireless environments. The required sensing time needed to achieve the desired probabilities of detection and false alarms in a low-SNR regime was measured. The minimum detectable signal strength due to the receiver noise uncertainties and background interference was also investigated. The experiments also measured the improvements in sensing performance obtained through network cooperation, identified the location- and time-relevant threshold rule for hard-decision combining, and quantified the effects of spatial separation between radios in indoor environments. In [415], the feasibility of cyclostationary feature detection has been investigated. It is shown through experiments that cyclostationary feature detectors require tight synchronization between the sampling clock and the signal of interest, so that the cyclostationary features can be useful in low-SNR regimes.

1.5.2 The Center for Wireless Telecommunications at Virginia Tech

A distributed genetic-algorithm-based CR engine is proposed in [367] [368]. The cognitive engine focuses on how to provide CR capability to the physical and MAC data link layers. The system is structured so that the cognitive engine can provide cognitive functionality that scales with primary users. Information about the radio spectrum environment and location of the users is used to better classify the environment and choose potential radio configurations by the engine. The cognitive system monitor

enables cross-layer cognition and adaptation by classifying the observed channel, matching channel behavior with operational goals, and passing the goals to a wireless system genetic-algorithm adaptive controller module to gradually optimize radio operation. The cognitive-engine framework is compared with the traditional adaptive-controller framework. It is shown that the cognitive engine can find the best tradeoff between a user's operational parameters in a changing environment while the traditional adaptive controller can only increase or decrease the data rate, wasting usable bandwidth or power due to its inability to learn.

Using this CR engine, an experiment was conducted in [289] to demonstrate the benefits of CR by dynamic spectrum sharing. The experiment focuses on the unlicensed 5.8-GHz ISM band to compare the spectrum utilization of IEEE 802.11 a/g physical layers with the CR version of such a WLAN radio. In the CR OFDM PHY layer model, the access point's channel is dynamically changed due to the location and interference level, and the subscribers can pick an access point according to the sensed SINR and load condition at each access point. By sensing the radio spectrum environment and making real-time decisions on frequency, bandwidth, and waveform, the CR OFDM PHY layer can achieve an increase of 20 dB in SINR over the standard OFDM PHY layer. A coexistence experiment conducted in [315] studied the feasibility of coexistence of the primary users and secondary users in a common spectrum band. In the worst-case scenario with no guard bands between the primary users and secondary users, the primary users can be minimally affected if the secondary users' transmissions are properly modified.

Issues involved in adapting CR technology to consumer markets are discussed in [21]. The pricing mechanism to trade the network resources will be integrated as part of the user domain and modeling system in the CR engine, and the pricing mechanism should hit a good balance between resource efficiency and computational complexity. It is worth studying the interaction of the proposed pricing system with the cognitive engine, and whether the CR engine is able to produce good solutions within a reasonable period of time.

1.5.3 WINLAB at Rutgers University

Researchers at Rutgers University have constructed an Open Access Research Testbed for Next-Generation Wireless Networks (ORBIT) [370] to perform experimentation on CR research. The ORBIT testbed has a two-tier architecture, consisting of an indoor radio grid emulator for controlled experimentation and an outdoor field-trial network for end-user evaluation in real-world settings.

Several of the key architectural issues for CR networks are discussed in [366], including spectrum agility and fast spectrum scanning over multiple frequency bands, fast PHY adaptation, the spectrum etiquette protocol and dynamic spectrum coordination, flexible MAC-layer protocols, control and management protocols, and ad hoc group formation and cross-layer adaptation. A high-performance CR platform with integrated physical- and network-layer capabilities [363] based on the architectural foundation [366] is under development using the ORBIT testbed. The CR

prototype's architecture consists of several major elements: an agile RF front-end working over a range of frequency, FPGA-based software-defined radio (SDR) to support a variety of modulation waveforms, a packet-processing engine for protocol and routing functionality, and an embedded CPU core for control and management. The goal of its hardware design is to provide fast RF scanning capability and the software will use the GNU software radio code base. This prototype is differentiated from other CR projects in that the design uses hardware accelerators to achieve programmability and high performance at each layer of the protocol stack.

An experimental study on spectrum sensing for localizing transmitters using sensor nodes with CR capability has been proposed in [365], where the sensor nodes can sense only a limited bandwidth at a time. Using triangulation techniques based on the detected power at each sensor, the experiments study how to localize a single transmitter and multiple asynchronous interfering transmitters that are transmitting in the same band, and how to find the spectral occupancy over a band of frequencies. It is shown through the experiments that energy-detection techniques are not sufficient to localize multiple transmitting sources.

1.5.4 Others

A real CR governed by a cognitive engine is proposed in [62], since most of the existing DSA protocols have been defined in such a way that they could not be directly implemented on a real CR. The cognitive engine provides the capability both to reason (i.e., AI planning) and to learn (i.e., machine learning). Reasoning helps decide the best action in a particular scenario given knowledge of how the actions will affect the progress toward an objective, while learning helps get more information about how a particular action will affect the overall system state by trying out the action. The work in [62] translates the basic semantics of DSA into the Action Description Language, and implements a primary-prioritized Markov spectrum-access algorithm [449] within the Open-Source Cognitive Radio, which allows spectrum sharing both in frequency and in time.

The secondary users can utilize the white space in the time domain by transmitting during the idle periods between primary users' packet transmissions. In order to make the best use of the white space, a realistic yet tractable model that can provide adequate prediction performance while achieving a balance between statistical accuracy and complexity needs to be established. An experimental testbed is developed in [151] to gather empirical data on the channel statistics. The testbed consists of a wireless router and several workstations with WLAN adapter cards. A vector signal analyzer is used to capture the raw complex baseband data, and various sensing strategies (energy- and feature-based detection) can be evaluated on the same data. This not only provides insight when developing real-time implementations, but also confirms the data validity.

A prototype of CR-based sensor implementation with off-the-shelf IEEE 802.11 devices was built in [388]. The sensor prototype uses WLAN cards with a built-in Atheros chipset, while slightly modifying the Atheros device driver to assess key ideas

of spectrum sensing. Important issues in spectrum sensing have been explored, such as how to choose the energy-detection threshold, how to characterize secondary traffic, and how to schedule the sensing priority. The experimental results provide guidelines for implementing a spectrum sensor in real CR networks.

The experimental CR research using commercial platforms is limited by their inability to provide full control of the RF, PHY, and MAC functionalities and change the underlying framework. The prototype system designed in [494] can provide more flexibility and reconfigurability. A real-time MIMO OFDM testbed was developed to support a large number of permutations of physical-layer modes, which are defined by the MAC through an API interface. The header of each MAC-to-PHY transmission contains the value of the configuration for that specific packet, and thus the higher layers can control the type of the packet and its operational mode. On the other hand, the PHY can provide SNR, CRC results, and channel state information to upper layers and enable advanced protocols. Besides the intelligent spectral-allocation feature, the testbed also supports a greater range of data rates and high throughput. The future version of the prototype is expected to allow independent allocation of the RF transceiver chains and intelligently determine the number of antennas used for transmission and sensing.

A CR testbed system employing a wideband multi-resolution spectrum-sensing (MRSS) technique is proposed in [184]. The testbed employs a vector signal generator and a vector signal analyzer to provide a variety of built-in-standard wireless signals, such as IEEE 802.16, IEEE 802.11 a/b/g, 3G-wireless, and digital video broadcasting. The hardware-control programs were developed in a Matlab environment. The received signal is investigated with the MRSS hardware to identify its spectral-occupancy status. Specifically, a wavelet transform is employed in the MRSS technique. By adapting the wavelet's pulse width and its carrier frequency, the spectral-usage status can be represented in multi-resolution format. The MRSS technique is shown to be able to examine a wideband spectrum and detect many sophisticated signal formats in current and emerging wireless standards. Therefore, the testbed can probably provide a flexible and versatile environment for developing CR access schemes.

The Kansas University Agile Radio (KUAR) platform is presented in [281] [282]. There is a very flexible RF front-end that can support a large center-frequency range as well as wide transmission bandwidths. The powerful on-board digital processing can support a variety of cognitive functions, such as implementing numerous modulation algorithms, MAC protocols, and adaptation mechanisms. The self-contained, small-form-factor radio unit enables convenient portability. Moreover, the KUAR platform is highly configurable in that it has a robust set of hardware and software tools that allow developers to work in their area without being encumbered by the other layers. The low-cost build cycle also facilitates broad distribution of the KUAR units to the CR research community.

A virtual-SDR system using an Atheros platform has been proposed in [105]. The experimental platform adopts a MultiMAC framework, which can dynamically reconfigure MAC- and physical-layer properties in order to achieve the best performance while providing the correct MAC-layer algorithm to decode the data frames.

Therefore, the platform can respond quickly to changes in the radio environment and requirements in optimizing the spectrum efficiency.

A software-defined CR prototype has been developed in [159], which consists of a hardware platform and a software platform. The hardware platform consists of a multi-band antenna with a frequency range in the UHF band and 2–5 GHz, a multi-band RF front-end, an FPGA-based signal-processing unit, and a CPU. The software platform is composed of several managers that control spectrum sensing and reconfiguration by changing software packages, where the software configuration specifies the type of communication system.

An adaptive wireless network testbed on the MIRAI Cognitive Radio Execution Framework is proposed in [201]. The physical layer accepts both a virtual and a real CR device with interfaces to the software environment through a gateway plug-in. Therefore, the testbed provides scalability for more than 10 000 nodes by combining real CR devices and virtual nodes. It enables the provision of a flexible configuration to any protocol and application, verifies protocols in the MAC layer, and allows for remote interaction with the testbed over the Internet. However, due to the processing-power limitation of a single PC and the overhead of the standard communications between the CR devices and the PC, the types of experiments that can be done on the testbed are limited and real-time signal processing becomes difficult.

1.5.5 Industry

Since 2005, the Shared Spectrum Company has been conducting field tests to measure the spectrum-occupancy status [278] in various locations, including outdoor urban and rural locations, and an indoor location. It has been found that there is significant spectrum white space, and an agile, dynamic spectrum-sharing (DSS) radio can provide high spectrum utilization. Motivated by observations from the measurements, the company started to design effective spectrum detectors and DSS radio. Two implementations of detectors with significantly different operating characteristics are studied in [398], which compares the probability of false alarms and the probability of detection, as well as analyzing how to obtain the threshold level of the detectors using data measured from the environments. Detection thresholds for safe operation in unoccupied TV bands without causing harmful interference with other authorized operations are examined in [298].

A policy-based network management framework for controlling the spectrum access is presented in [331], including a prototype implementation and demonstration. This approach can support easy reconfiguration and policy authoring, secure policy distribution, management and enforcement, automated policy synchronization, conflict resolution, and opportunity discovery. A field framework experimentation is presented in [332], where the distributed, policy-driven system restricts spectrum access on the basis of spectral, temporal, and spatial context, while fully utilizing the available spectrum compared with traditional static spectrum access.

The MITRE corporation has developed a testbed, the adaptive spectrum radio (ASR) [180], to demonstrate the feasibility of the ASR concept. The ASR is expected to