University of Alberta

OPTIMAL OPPORTUNISTIC CHANNEL ACCESS IN WIRELESS COMMUNICATION NETWORKS

by

Zhou Zhang

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Communications

Department of Electrical and Computer Engineering

Fall 2013
Edmonton, Alberta

Abstract

High transmission rates will be demanded in future wireless communication networks. However, we will soon experience a spectrum scarcity problem since almost all available spectrum has been allocated to various wireless applications. A promising solution to this problem is to significantly improve the spectrum utilization efficiency by using opportunistic channel access (OCA). In the literature, OCA approaches have been developed in two kinds of networks: cognitive radio networks (CRNs) (in which secondary users, which are unlicensed users, may access the spectrum when primary users, which are licensed users, are not active) and wireless networks exploiting time diversity (in which a user may give up its transmission opportunity if its channel is deeply faded). This thesis is focused on optimal OCA in such two networks, with four research components.

The first three research components are to achieve optimality in CRNs. The first research component is for the scenario that the statistical information of primary traffic (such as busy/idle probabilities) is known at a secondary user. When a secondary user can sense multiple channels simultaneously but the maximum numbers of channels that can be sensed simultaneously and that can be accessed simultaneously are both limited, we derive optimal strategies to select which channels to sense and which sensed-free channels to access. The second research component is for the scenario that statistical information of primary traffic is unknown and thus needs to be learned during channel sensing and access process (which results in learning loss). When busy/idle states of each channel are independent from one

slot to another, we derive secondary channel sensing and access rules with asymptotically finite learning loss or logarithmic learning loss. As an extension of the second research component, the third research component uses another popular channel statistical behavior model: busy/idle states of each channel over time slots follow a Markov chain. We derive a channel sensing and access rule with logarithmic learning loss.

The last (but not the least) research component is for optimal distributed OCA that utilizes time-diversity in wireless cooperative networks. Two cases are considered: the case when the source knows channel state information of links from itself to relays and from relays to its destination; and the case when a source knows only channel state information of links from itself to relays. In the two cases, the optimal transmission strategies that maximize the average system throughput are derived theoretically. Our research reveals that time diversity can be exploited in a wireless cooperative network by our proposed strategies.

Acknowledgments

I sincerely appreciate my supervisor, Dr. Hai Jiang, for his continuous support and patient guidance during my PhD program. His devotion and expertise in research motivated me. His timely and edifying feedback as well as his encouragements inspired me and helped me find the path to success in my research. His precision, his rigorous attitude, and his high standard at work have set an example for me in the future career.

I would like to express my gratitude to the members of my defense committee, Dr. Yindi Jing, Dr. Vincent Wong, Dr. Di Niu and Dr. Majid Khabbazian, for their efforts reviewing my thesis and providing comments on my dissertation. Their feedback helped me improve the quality of my thesis.

Many thanks to my current and previous colleagues: Jie Gao, Lijie Huang, Qian Cao, Saman Atapattu, Qiang Guo, Qian Wang, Yichen Hao and Lei Dong, for their inspiring discussions. The friendship with them leaves me many pleasant recollections of the four years of my life in Canada.

Most importantly, I am very grateful to my family including my grandparents, parents, parents-in-law, my wife and little daughter, Yuexin Zhang. Their support raised me up in all the frustrations during my program. A special gratitude goes to my wife, Ting Zhang who accompanied me experiencing both happiness and difficulties in the past years.

Finally, I would also like to thank all my teachers, friends and relatives. Without their support and encouragements, the current success could not be achieved.

Last but not the least, I greatly appreciate the financial support from China Scholarship Council during the whole four years of my doctor program and Alberta Innovates Technology Futures during the fourth year of my program.

List of Symbols

$a.s.$	Almost surely
B	Transmitted bits in the data transmission period
$\mathbb{E}[\cdot]$	Expectation operator
$\mathbb{E}_1[\cdot]$	Expectation operator on the main layer of stopping problem
$\mathbb{E}_2[\cdot]$	Expectation operator on the sub-layer of stopping problem
f_{ij}	Channel gain from the ith source to the jth relay
$f(\theta_i)$	Probability of Channel i being sensed free
g_{ji}	Channel gain from the jth relay to the ith destination
$\mathbb{I}[\cdot]$	Indicator function
$\mathcal{I}_{\mathcal{M}}$	Set of channels sensed free if channels in $\mathcal{M}$ are sensed
$\mathcal{I}_{\mathcal{M}}(j)$	Set of channels sensed free if channels in $\mathcal{M}$ are sensed at Slot j
$\mathcal{I}(j)$	Set of channels sensed idle at Slot j
$\mathcal{K}$	Set of channels to access
$\mathcal{K}(j)$	Set of channels to access at Slot j
L	Number of relays
$\mathcal{M}(j)$	Set of channels to sense at Slot j
$\mathcal{M}^*$	Optimal set of channels to sense
$\mathbb{P}(\cdot)$	Probability operator
P_d^i	Detection probability for sensing Channel i
P_f^i	False alarm probability for sensing Channel i
P_r	Transmission power of a relay
P_s	Transmission power of a source
p_{00}^i	Transition probability of Channel i from a busy state to a busy state
p_{11}^i	Transition probability of Channel i from an idle state to an idle state
p_0	Channel contention probability of each source

p_1	Channel contention probability of each relay
$R(t, \psi)$	Regret of rule ψ until Slot t
R_n	Achievable rate in the nth observation via the best relay
R_m	Achievable rate in the mth observation at second hop
$S_i(j)$	Idle/Busy state of Channel i at Slot j
$s(n)$	Winner source of channel contention in the first-hop observation n
$s(m)$	Winner relay of channel contention in the second-hop observation m
T	Duration of a time slot in cognitive radio networks
t_l	Time spent in the lth observation
$X_i(j)$	The sensing observation of Channel i at Slot j
$\mathbf{X}(t)$	Sensing observation of all channels at Slot t
$\lvert \cdot \rvert$	Cardinality of a set
$\lfloor \cdot \rfloor$	Floor function
$\lceil \cdot \rceil$	Ceiling function
θ_i	Idle probability of Channel i
$\hat{\theta}_i$	Estimated idle probability of Channel i
$\boldsymbol{\Theta}$	Idle probability vector of all channels
$\hat{\boldsymbol{\Theta}}$	Estimation of $\boldsymbol{\Theta}$
σ_f^2	Variance of f_{ij}
σ_g^2	Variance of g_{ji}
δ	Duration of a time slot used in channel contention in AF relay networks
τ	Sensing period of a slot
τ_{RTS}	Duration of an RTS
τ_{CTS}	Duration of a CTS
τ_d	Data transmission time
τ_o	Mean of the duration of an observation
τ_o^s	Mean of duration of an observation in the main layer
τ_o^r	Mean of duration of an observation in the sub-layer
λ^*	Maximal system throughput

List of Abbreviations

Acronyms	Definition
ACK	acknowledgement
AF	amplify and forward
AIFS	arbitration inter-frame space
BER	bit error rate
CDF	cumulative distribution function
CDMA	code division multiple access
CRN	cognitive radio network
CSI	channel state information
CTS	clear-to-send
CSMA	carrier sense multiple access
CSMA-CD	CSMA with collision detection
CSMA-CA	CSMA with collision avoidance
DF	decode and forward
FCC	Federal Communications Commission
FDMA	frequency-division multiple access
i.i.d.	independent and identically distributed
LTE-A	advanced long term evolution
MABP	multi-armed bandit problem
MAC	medium access control
MIMO	multiple-input-multiple-output
OCA	opportunistic channel access
OCSA	opportunistic channel sensing and access
RTS	request-to-send

SNR	signal-to-noise ratio
TDMA	time-division multiple access
WLAN	wireless local area network
WRAN	wireless regional area network
UCB	upper confidence bound

List of Tables

List of Figures

Table of Contents

Chapter 1

Introduction

With the rapid development of wireless communication technology in the past decades, very high speed data transmission is expected to be demanded, to provide wireless Internet service anywhere and anytime for emerging wireless multimedia applications. Increasing data traffic demand has led to significant expansion of wireless networks in the past, but also resulted in shortage of communication resources (bandwidth resources, buffer resources, computation resources, etc.).

To meet the high-speed service requirements for existing and emerging wireless applications in the near future, one possible solution is to use new spectrum with large bandwidth. However, we will soon experience a spectrum scarcity problem since almost all available spectrum has been allocated to various wireless applications. Therefore, to achieve high data rate, we have to significantly improve spectrum utilization efficiency. Three main challenges that affect spectrum utilization efficiency are as follows.

First of all, in the current spectrum regulation structure in different countries, only licensed users have the permission to access a specific spectrum band. On the other hand, the licensed spectrum is severely under-utilized, as evidenced by an actual spectrum usage measurement conducted by the US Federal Communications Commission (FCC) Spectrum Policy Task Force [1] that shows that, at any particular time, a large portion of licensed spectrum is actually not utilized by licensed users. This implies that the spectrum scarcity problem is actually due to spectrum under-utilization rather than the lack of new spectrum.

Secondly, the wireless channel fading largely affects spectrum utilization efficiency. In wireless transmissions, the transmitted signal normally suffers from path loss, large-scale fading and small-scale fading. The path loss means the reduction of the signal's power in the order of d^{κ}, with d being the distance between the transmitter and the receiver and κ being path loss attenuation exponent (varying from 2 to 4 in different environments). The large-scale fading, also known as shadowing or slow fading, is due to the blocking effect of some objects (such as mountains, high buildings, etc.) between the transmitter and the receiver. Small-scale fading, also known as multipath fading or fast fading, arises from the constructive or destructive effect of signals from multiple paths (due to reflection, scattering, or diffraction by objects) between the transmitter and receiver.

Thirdly, multiple access largely affects spectrum utilization efficiency. The wireless transmission medium is usually shared by multiple users. When multiple users in a neighborhood transmit simultaneously over the same spectrum band, they generate large interference to each other, and their packets may even collide with each other. Thus, the capacity obtained by each user is heavily degraded. So far this challenge has been well addressed in centralized networks (e.g., cellular networks) where a central controller (e.g., a base station) helps schedule the transmissions from multiple users. However, it is much challenging to achieve high spectrum efficiency in a distributed network, and it is harder to achieve spectrum utilization optimality.

To address the challenges in spectrum efficiency, three promising solutions, namely *cognitive radio*, *cooperative diversity* and *time diversity*, have been introduced in the literature and have attracted tremendous attention recently.

1.1 Cognitive Radio

The cognitive radio, which was originally introduced in [2], was considered to have great potential to alleviate the scarcity of available spectrum and the under-utilization of licensed spectrum, since it enables efficient spectrum utilization in an opportunistic manner. In cognitive radio networks (CRNs), nodes that have licence to operate in a spectrum band are referred to as *primary users*, while nodes that do not have license are called *secondary users*. Secondary users can utilize the spectrum opportunistically, i.e., in a non-intrusive manner. In such a channel access structure, primary users always have priority to access the spectrum, while secondary users are required not to affect primary users' activities whenever possible. Based on cognitive radio technique, some wireless communication standards are under development [3], [4], such as the IEEE 802.22 wireless regional area network (WRAN) standard that provides broadband Internet access in rural and remote areas.

Secondary users can use either of two approaches to share the licensed spectrum: *spectrum underlay* approach and *spectrum overlay* approach.

With the spectrum underlay approach, secondary users are permitted to use the licensed spectrum at any time. However, their generated interference at the primary users' sites should be below a tolerable threshold such that primary users can still have reliable transmissions. This means that the power spectrum density of signals of secondary users is very low. To meet secondary users' quality-of-service (QoS) requirement, secondary users may spread their signals over a large bandwidth for transmission, for example, using ultra-wideband communications or using code division multiple access (CDMA) technique to spread the spectrum.

In the spectrum overlay approach, secondary users are allowed to use *spectrum holes* (which are the spectrum bands that are not being used by primary users at a time period) to opportunistically transmit their data. The limit of power spectrum density in the spectrum underlay approach does not apply in the spectrum overlay approach. Thus, the spectrum overlay approach has attracted lots of research attentions recently [5]–[17]. Since primary users always have channel access priority, secondary users in the spectrum overlay approach need to first check (by using spectrum sensing techniques such as energy detection [18], matched filter detection [19], cyclostationary feature detection [20], etc.) whether the channel is busy (i.e., being occupied by primary users) or idle. If the channel is sensed idle, the secondary users can access the channel for a limited time period, and then start spectrum sensing again; otherwise, the secondary users wait for a time period and start spectrum sensing again. During data transmission, secondary users may also need to keep monitoring primary activities. If primary users are detected to be back, secondary users are required to stop transmission over the channel immediately, and may switch to other idle channels to keep service continuation. In other words, opportunistic channel access (OCA) is adopted by secondary users.

1.2 Cooperative Diversity

An effective approach to overcome negative effects due to channel fading is to utilize the spatial diversity [21] in a system that is equipped with multiple sufficiently-spaced antennas that transmit/receive the same information, by using, for example, the multiple-input-multiple-output (MIMO) technique. However, in a typical wireless network, mobile user devices usually have a small size. Therefore, it is very hard to deploy multiple antennas on a mobile device. To solve this problem, a new paradigm called *cooperative transmission* was introduced, to achieve *cooperation diversity* (a new form of spatial diversity) [22], [23].

In cooperative transmission, the advantage of spatial diversity can be achieved by letting one or more intermediate relaying nodes forward information from the source to the destination. With the relaying node(s), a virtual multiple-antenna system is built up. Since the signal of the source has multiple paths to the destination (direct path and relay paths), the probability that the destination receives an overall heavily-degraded signal can be largely reduced, which means the negative effect of channel fading can be significantly alleviated. Cooperative transmission can improve both system throughput and energy consumption, with a higher diversity order [24]–[26].

Two popular schemes are adopted in cooperative communications, namely *amplify and forward* (AF) and *decode and forward* (DF) [24], [26]. The AF scheme is the simpler one, in which a relay node forwards to the destination an amplified version of its received signal from the source without decoding the message. The drawback is that the noise at the relay is also amplified, referred to as *noise propagation*. In the DF scheme, a relay node first decodes the received signal from the source. It then re-encodes the message either using the same code-book as used in the message from the source (called regenerative DF) or adopting an independent code-book (called non-regenerative DF). The relay node then forwards to the destination the encoded signal.

1.3 Time Diversity

Apart from cooperative diversity, time diversity is another new form of diversity that also has the potential to significantly increase the spectrum utilization efficiency. The major idea is that, for transmission of a source to its destination, since the channel varies with time (due to fading), it is desired that the source transmits only when its channel is good such that a high transmission rate can be achieved. This idea was first explored in centralized wireless networks (for example, a cellular network) in which a central node such as the base station coordinates the transmission schedules of multiple users in the medium access control (MAC) layer. To exploit the time diversity, the MAC layer protocol should be jointly designed with the physical layer, thus leading to a cross-layer design concept, namely channel-aware scheduling or OCA. In specific, if a transmitter is found to have poor channel quality, the central node may ask the transmitter to give up its channel access opportunity to other users that have good channel quality (since those other users can gain more by utilizing the channel). In a long term, all users will benefit, since although a user may sacrifice its channel access opportunity when its channel is poor, it can get much more when its channel is better by utilizing the channel access opportunity of other users that have poor channel quality. It can be seen that it is not difficult to achieve optimal OCA that maximizes the time-diversity gain in a centralized network [27], [28], in which the central controller is responsible to collect the channel state information (CSI) of all users and schedule only those users with the best channel conditions. On the other hand, the research on distributed OCA is still in its infancy. Without a central controller, it is hard for a user to decide when to give up its transmission opportunities. An intuitive way is to categorize the channel of a user into two states: *good state* when the channel gain is above an arbitrarily selected threshold; and *bad state* otherwise. Then a user gives up its channel access opportunity when its channel is bad. Apparently time diversity (i.e., a user experiences different channel gain when time varies) is not fully utilized by the intuitive method.

It is worth noting that the OCA in CRNs is in a reactive manner (i.e., secondary users *wait* for transmission opportunities), while the OCA that exploits time diversity is in a proactive manner (i.e., a user *offers* its transmission opportunity to others).

1.4 Thesis Motivations and Contributions

This thesis focuses on optimal OCA in both CRNs and time-diversity approaches.

1.4.1 Optimal Opportunistic Channel Sensing and Access in CRNs

In CRNs, since the spectrum overlay approach is much more popular than the spectrum underlay approach, it is adopted in this thesis research. In the spectrum overlay approach, a secondary user needs to first sense the channel, and can access the channel only when the channel is sensed free (which means primary users are idle). Therefore, a slotted time structure is usually adopted: time is partitioned into slots, and each slot consists of a sensing period (used by secondary users to sense the channels) and a transmission period (used by secondary users to transmit over sensed-free channels). For such a slotted structure, OCA is normally jointly designed with channel sensing strategies (such as strategies that decide which channels are selected to sense in the sensing period of a time slot), referred to as *opportunistic channel sensing and access (OCSA)* in the sequel.[1] Optimal OCSA has been well investigated in the literature [8], [11]–[15], [29]–[31]. Most of the research efforts assume that the channel sensing is perfect with no missed detections or false alarms. However, in a practical CRN, the channel sensing is generally imperfect. To fill this research gap, optimal OCSA with imperfect channel sensing is investigated in this thesis in two directions: when the statistical information of primary traffic is known and unknown, as follows.

By assuming that the statistical information of primary traffic (for example, the availability information of the channels) is known at a secondary user, we are interested in two questions: which channels should be sensed in the sensing period of a time slot? and which sensed-free channels should be accessed in the transmission period of the time slot? To be more specific, a secondary user usually has a large number of potential channels, but cannot sense all of them simultaneously at the sensing period. Then it needs to select a subset of the channels to sense. Due to hardware or power constraints, the secondary user may not be able to access all sensed-free channels in the transmission period. Therefore, it needs to select a subset of the sensed-free channels to access. These two questions are answered in Chapter 3 of this thesis, which derives optimal OCSA strategies that maximize the secondary user's average reward at a slot.

By assuming that the statistical information of primary traffic is unknown at a sec-

[1]To avoid confusion, in the sequel of the thesis, OCA only means opportunistic channel access exploiting time diversity, while OCSA means opportunistic channel access in CRNs that is jointly designed with channel sensing.

ondary user, a secondary user needs to learn the information in its channel sensing and access process, and thus, learning loss (also called *regret*) is inevitable. We are still interested in the two questions: which channels to sense and which channels to access? Our target is the minimal regret (compared to the genie-aided ideal case that knows the statistical information of primary traffic). Although the problem was solved in the literature when the channel sensing is perfect, the existing solutions/methods do not work anymore when channel sensing is imperfect. To fill the research gap, we solve the problem with imperfect channel sensing, in two popular channel availability models: 1) for each channel, the busy/idle states at different slots are independent and identically distributed (i.i.d.), referred to as *i.i.d. model*; 2) the busy/idle states of each channel over time slots follow a Markov chain, referred to as *Markov model*. The optimal OCSA with the i.i.d. model is addressed in Chapter 4. Since regret is inevitable, the most desired OCSA rules are those with asymptotically finite regret. When all potential channels can be sensed simultaneously, we derive OCSA rules with asymptotically finite regret, which are most desired. On the other hand, when the secondary user cannot sense all channels simultaneously, we prove that the best possible OCSA rules are the *order optimal* rules (i.e., rules with logarithmic regret). We also derive order optimal OCSA rules. In Chapter 5, an OCSA rule is proposed and proved to be order optimal for the case that the busy/idle states of each channel follow the Markov model.

1.4.2 Optimal Distributed OCA in Wireless Cooperative Networks

It is well recognized that, by exploiting the cooperative diversity, wireless cooperative networks can largely improve the spectrum efficiency. On the other hand, by exploiting time diversity, OCA can also largely enhance the system throughput. OCA has been well studies in centralized networks [27], [32], and has also be investigated recently in a distributed network [33]–[35]. Then a natural question is: can we (and how to) further exploit time diversity in a distributed wireless cooperative network? This question is answered in Chapter 6 in this thesis. A wireless cooperative network with multiple source-destination pairs and multiple AF relays is considered. All source nodes contend through a random access procedure. For the case that a source has CSI of links from itself to relays and from relays to its destination and the case that a source has only CSI of links from itself to relays, we theoretically prove the existence of optimal OCA strategy, and derive optimal strategies for the two cases. The derived optimal OCA strategies are with simple structure, and thus, can be easily implemented in a practical distributed network. Our research reveals that time diversity can also be exploited in a wireless cooperative network by our derived strategies.

1.5 Thesis Outline

The thesis is organized as follows. In Chapter 2, basic concepts of OCA and OCSA are introduced, and related works are surveyed. Optimal OCSA strategies when statistical information of primary traffic is known are presented in Chapter 3. Optimal or order optimal OCSA strategies without statistical information of primary traffic are proposed and theoretically analyzed in Chapter 4 and Chapter 5, for the i.i.d. model and the Markov model, respectively, of busy/idle states of a channel over time slots. Distributed OCA in wireless AF relay networks is discussed in Chapter 6. Chapter 7 concludes the thesis, and gives possible future research directions as well.

Chapter 2

Background and Literature Review

The wireless medium is a shared medium by a number of wireless users, and thus, the access to the medium by all the users should be coordinated effectively and efficiently. In general, the wireless channel access can be either *horizontal* or *vertical* [37]. In horizontal channel access, all the users have the same priority, for example, in a traditional wireless local area network (WLAN) with the same type of users. In vertical channel access, some users are given higher priority than other users, and thus can have an advantageous position in channel access. One example is the IEEE 802.11e-based WLANs, in which users with shorter arbitration inter-frame space (AIFS) and smaller backoff window can gain more channel access opportunities [38]. Another example is the CRNs, in which primary users should always have higher priority than secondary users.

2.1 OCA in Wireless Networks

2.1.1 Traditional Wireless Channel Access

Wireless channel access schemes can be categorized into two classes: centralized channel access and distributed channel access.

In a centralized channel access, a central controller, such as the base station in a cellular network and the access point in a WLAN, is responsible to decide how the resources (frequency, time, code channels, etc.) are shared among the users and how much power level a user can use. In a fixed channel assignment, each user is assigned a fixed frequency band in frequency-division multiple access (FDMA), a fixed time slot in time-division multiple access (TDMA), or a fixed code channel in code-division multiple access (CDMA). Such fixed channel assignment is easy to implement since it needs to be done only when a call is admitted. Its major drawback is potential waste since the assigned resources are exclusively used by the particular users. To improve the resource utilization, dynamic channel access lets all resources shared by all users [39], [40]. The central controller first collects information (for example, queueing size at each user, and waiting time of each packet in the queue) from all users at each time slot and then decides when and how a user can transmit. So the major drawback of dynamic channel access is the communication overhead and computation complexity.

In a distributed channel access (also known as random channel access), no central controller is involved in the channel access, and thus, the users need to decide by themselves whether to transmit and if yes, how to transmit (e.g., with what power level?) at each time slot, based on their own local observations of the network. The first distributed channel access scheme is the ALOHA [41], in which each user transmits whenever it has traffic. When two transmissions have overlap in time, then the two transmissions collide with each other. Therefore, the normalized throughput of ALOHA is low (18%). The normalized throughput is doubled if all transmissions are synchronized at the beginning of a slot, referred to as slotted ALOHA [42]. Some other variants of ALOHA can be found in [43], [44]. On the other hand, in the literature, it was also proposed that each user senses the channel first and transmits if the channel is idle, or defers otherwise. This is the basic idea of carrier sense multiple access (CSMA) and its variants: CSMA with collision detection (CSMA-CD) used in Ethernet, and CSMA with collision avoidance (CSMA-CA), which is the basis for the IEEE 802.11 Medium Access Control Standard [45].

2.1.2 Centralized OCA

To exploit time diversity in a wireless network, it is desired that only users with the best channel quality can access the channel. This can be achieved in a centralized channel access, by letting the central control collect CSI from the users, select those users with the best channels, and announce the scheduling decision to all the users. Therefore, centralized OCA has been well studied in cellular networks, which are typical centralized networks with the base station working as the central controller.

Centralized OCA strategies can be found, for example, in [46]–[50], for uplink transmission in cellular networks. While most of existing OCA efforts assume that CSI is known at the scheduler and that each user has infinite data to transmit, some other works also investigate channel-aware queue-aware rules taking into account queue state of each user [51]. *Ergodic scheduling* is investigated in [52], [53], which maximizes the expected system throughput subject to imperfect CSI at the scheduler.

For downlink transmission in cellular networks, two major classes of OCA schemes are: *margin adaptive* (which satisfies user QoS requirement) and *rate adaptive* (which maximizes system throughput) [54]. In [55]–[57], margin adaptive OCA schemes are presented to minimize the total transmit power while satisfying each user's QoS requirement, for example, in terms of data rate and bit error rate (BER). In [58]–[66], rate adaptive schemes are presented that maximize the system throughput constrained by the total transmit power. All these works assume perfect CSI is available at the scheduler. By assuming imperfect CSI at the scheduler, ergodic scheduling is investigated in the downlink of cellular networks in [60], [67], which maximizes the expected system throughput subject to imperfect CSI.

2.1.3 Distributed OCA

In centralized OCA models, the central controller is responsible to collect CSI from all users and conduct channel/power assignment. Therefore, the communication and computation overhead may be large. In some cases, communication overhead to get instantaneous CSI may not be tolerable. In other cases, the central controller may not exist (for example, in a self-organized ad hoc network). Therefore, OCA in a distributed manner, called distributed OCA, is more interesting and challenging.

In distributed networks where multiple users randomly access the channel, it is hard for a user to decide when it is optimal to give up its transmission opportunities. An intuitive way is to categorize the channel of a user into two states: good state when the channel gain is above an arbitrarily selected threshold; and bad state otherwise. Then a user gives up its channel access opportunity when its channel is bad. Apparently time diversity (i.e., a user experiences different channel gain when time varies) is not fully utilized by the intuitive method. This problem was addressed recently in [33], by means of optimal stopping. The major idea is to let all the users contend for channel access. It is found that, 1) if the winner in a contention has an achievable (transmission) rate smaller than a threshold (which can be obtained numerically), it is optimal for the winner to give up its transmission opportunity and all users continue to contend; and 2) if the winner in a contention has an achievable rate larger than the threshold, it is optimal for the winner to *stop* here, i.e., to utilize the transmission opportunity and transmit its data. The beautiful part of the work is in the *pure-threshold strategy*, which is easy to implement. As extensions to the work in [33], interference channel which can tolerate multiple users transmitting is considered in [34] where more than one node can share the channel simultaneously, and delay constraints are considered for real-time service in [35]. Pure-threshold strategies are also derived in [34], [35].

2.1.3.1 Optimal Stopping Problem Background

The following is a brief introduction to the optimal stopping problem. Please refer to [36] for detailed discussion.

The theory of optimal stopping is to decide the best time to stop observing a random process. The random process is sequentially observed. The player has some knowledge of the random process (such as the distribution of the observed value at any moment). At any moment, after the observation of the random process is obtained, a player needs to decide whether 1) to stop the observation, take a pre-defined action, and get the reward, or 2) to continue to observe the random process. The target is to maximize the long-term reward. For example, consider a seller who would like to sell a house. The seller receives an offer every day. Denote X_n as the offer received on the nth day. The seller needs to decide whether to accept the offer (i.e., to stop observation) and sell the house (the pre-defined action) or to decline the offer and wait for other offers later. If the seller accepts the offer, the net reward is $X_n - nC$, in which C is the daily cost to keep the house (such as property tax and fees for electricity, water, gas, and waste management). It is desired that optimal stopping rule is taken to maximize the net reward.

To mathematically model a stopping problem, we can denote the observations at different moments as $X_1, X_2, ..., X_n$, where $1, 2, ..., n$ are the time indexes. Define Y_n as the net reward if the player stops after the nth observation. Y_n is a random variable. In the house-selling example, $Y_n = X_n - nC$. So at moment n, if the player stops, then he/she gets reward Y_n; otherwise, the player continues to observe X_{n+1} and the player's expected reward is $\mathbb{E}[Y_{n+1}]$, in which $\mathbb{E}[\cdot]$ means expectation.

To solve an optimal stopping problem, we must first prove that, there exists an optimal stopping rule. The existence of an optimal stopping rule can guarantee that the supremum of all expected rewards from any stopping rule can be attained. Two conditions that guarantee existence of an optimal stopping rule are:

$$\text{Condition 1.} \mathbb{E}[\sup_{n} Y_n] < \infty \tag{2.1}$$

$$\text{Condition 2.} \lim_{n\to\infty} \sup Y_n \leq Y_\infty \; a.s. \tag{2.2}$$

Here Y_∞ defines the reward if the player never stops, while "$a.s.$" is the short form for the terminology "almost surely", which means that Condition 2 is satisfied in a sense of probability 1.

Two most challenging tasks in solving an optimal stopping problem are the proof of existence of an optimal stopping rule and the derivation of an optimal stopping rule.

In some optimal stopping problems, the target is to maximize the average reward per unit of time. A typical example is the wireless communication system considered in [33]–[35]. A number of wireless users share a wireless channel. So the users contend for channel access. When a user succeeds in the channel contention, called a winner user, it observes its channel gain, and decides whether to stop (i.e., transmit over the channel within a fixed transmission time by using a transmission rate determined by its channel gain) or give up this transmission opportunity (i.e., all users contend again for the channel access and subsequently the winner user observes its channel gain). If a winner user stops, the reward is the amount of information bits that can be transmitted. The target is to maximize the average reward per unit of time, which is also the average system throughput. The optimal stopping problem can be formulated as: to find an optimal stopping rule whose average reward per unit of time is $\sup_{N>0} \frac{\mathbb{E}[Y_N]}{\mathbb{E}[T_N]}$, in which N denotes the stopping time, Y_N is the reward (i.e., the amount of information bits transmitted), T_i is the time duration until the ith observation, and T_N is the time duration until a stop.

Generally it is difficult to find an optimal rule that can maximize $\frac{\mathbb{E}[Y_N]}{\mathbb{E}[T_N]}$. However, the problem can be transformed into a classical form of optimal stopping problem, as follows. In the transformation, a price $\lambda > 0$ is introduced for the time cost T_N. Then a new stopping problem is formulated, which maximizes $\mathbb{E}[Y_N - \lambda T_N]$. Let N also denote stopping rule. If an optimal stopping rule for the new stopping problem can be derived for any λ, denoted as $N(\lambda)$, and if there exists a special price value λ^* such that $\mathbb{E}[Y_{N(\lambda^*)} - \lambda^* T_{N(\lambda^*)}] =$

$\sup_{N>0} \mathbb{E}[Y_N - \lambda^* T_N] = 0$, then it can be proved that $N(\lambda^*)$ is an optimal stopping rule for the original problem. Two challenging issues are 1) to prove existence of optimal stopping rule and derive an optimal stopping rule for the new stopping problem with any λ, and 2) to derive the special price value λ^*.

2.2 Channel Access through Cooperative Networks

In a wireless cooperative network, the users with different and varying channel conditions cooperate with each other and relay each other's packet to the destination. Through cooperation, cooperative diversity is exploited. Cooperative communication also has the capability to enlarge network coverage.

Thanks to the advantages, academia and wireless industry have been working on relay-based architectures in various cooperative networking environments. For example, although early WLAN standards (such as IEEE 802.11a/b/g) do not support relay-based transmission due to the nature of channel contention, recently a number of medium access control protocols have been proposed and investigated to support relays [68]–[73], and WLANs that support relays are amended in IEEE 802.11s [74].

For standards in cellular networks, relaying technique is not supported in early IEEE 802.16 standard until IEEE 802.16e. Following IEEE 802.16e, the IEEE 802.16j supports multiple-hop relaying. In addition, the 3GPP advanced long term evolution (LTE-A) is evolving to support relaying technique [75], [76].

The cooperative communications can also be very helpful in other networks when spectrum bandwidth and energy are both limited. For example, consider a wireless sensor network in which each sensor node is with low cost and limited energy supply. The dense deployment of sensors may degrade the spectrum efficiency due to the large interference generated by sensors. In this situation, cooperative communications can help to largely improve communication efficiency, as discussed in [77]–[80].

2.3 Optimal OCSA in CRNs

As a novel idea of vertical spectrum sharing, cognitive radio communications can largely enhance the spectrum efficiency by utilizing the spectrum holes. As an intelligent system, the CRN has the capability to learn the networking environment and adapt its transmission accordingly. To be more specific, a secondary user first senses the primary channels before its channel access, and during its transmission, the secondary user should immediately stop transmission over the channel if the primary users are back. In the literature, OCSA in CRNs can be categorized into two classes: OCSA with and without statistical information of primary traffic.

2.3.1 Optimal OCSA with Statistical Information of Primary Traffic

Optimal OCSA when secondary users have statistical information of primary traffic, such as information of available probabilities of primary channels, has been extensively studied in the literature [6]–[8], [11], [13]–[15], [81] to maximize the spectrum utilization efficiency.

Optimal Sensing Time Setting: In a CRN, the sensing time duration is essential, since a longer sensing time can more reliably detect primary activities while a shorter sensing time means more time for secondary transmission. The optimal sensing time setting is addressed in [8] for a CRN with a single channel. Time is partitioned into slots, and each time slot has a sensing portion and a transmission portion. In each slot, the target channel is first sensed in the sensing portion, and if it is sensed free, the secondary user transmits in the transmission portion. Optimal duration of the sensing portion is derived that maximizes the throughput of the secondary user. The optimal channel sensing time for multiple-channel CRNs is derived in [11] with slotted time sensing mode (i.e., the sensing time of a channel is one or more minislots) and continuous time sensing mode (i.e., the sensing time of a channel can be any arbitrary continuous value within the sensing portion of a time slot).

Optimal Sensing Order: In some CRNs, a secondary user may be designed to sequentially sense the potential channels one after another until a free channel or a free channel with good quality is found. The optimal sensing order in the sequential sensing is derived in [7] for a single user case. It is shown that, if adaptive modulation is not used, then it is optimal to have a sensing order being the descending order of the channel available probabilities, which is intuitive. However, if adaptive modulation is used, the intuitive sensing order may not be optimal in general. And optimal sensing order in this case is derived. The sensing order setting for a two-user case is solved in [82]. Two low-complexity suboptimal schemes are proposed, whose performance are shown to be very close to the performance of the exhaustively-searched optimal scheme.

Asynchronous OCSA: Since the slotted structure may require time synchronization with primary users, an asynchronous OCSA is studied in [81]. A secondary user can start sensing at any arbitrary moment, and access the channel if the channel is sensed free. In an ideal case with perfect spectrum sensing, an optimal OCSA policy is developed that can maximize secondary user throughput. In a practical case with imperfect spectrum sensing, a modified threshold-based OCSA is presented, which can obtain near-optimal performance.

Optimal Power Allocation: In [13], sensing time and power allocation are jointly optimized to maximize the secondary system throughput in a multiple-channel spectrum-overlay CRN, conditioned on that the detection probability of primary activities is above a certain threshold. In [14], optimal power allocation that maximizes the secondary system throughput is derived in a single-channel spectrum-underlay CRN, in which the secondary users' transmit power has either short-term or long-term constraint and the interference level to primary users also has either short-term or long-term constraint.

In addition, since cognitive radio has been a popular research area recently, several surveys/reviews are available that discuss some related issues in CRNs. Reference [83] gives an overview for important issues for OCSA in CRNs, such as network architecture, spectrum sensing, spectrum sharing, and spectrum mobility. Some main features of multiple-channel MAC protocols in CRNs are discussed in [84]. The main differences of cognitive radio MAC protocols from those in traditional networks are highlighted in [85]. In [86], control channel implementation for OCSA in CRNs is categorized in four classes: common control channel, hopping control channel, split phase control channel, and multiple rendezvous control channel. And a comparison is conducted among the four classes in terms of system throughput.

2.3.2 Optimal OCSA without Statistical Information of Primary Traffic

Research on the optimal OCSA without *a priori* statistical knowledge of primary traffic is in its infancy. Since the statistical information of primary traffic is unknown, secondary users need to learn the information during its channel sensing and access process. For example, by sensing a channel in a large number of different time slots, the available probability of the channel can be estimated, referred to as *channel exploration*. However, if the channel is with a low available probability, the sensing of the channel only brings few rewards. This is because, since the sensing results of the channel may often be "busy", the secondary user cannot access the channel after the sensing. In other words, the secondary user may need to take a long time to sense those not-good channels before the secondary user realizes that those channels are not good. This means that learning loss (also called regret) is inevitable, compared to the genie-aided ideal case that the secondary user knows statistical information of primary traffic and thus can always select the optimal channel(s). To minimize the regret, it is essential to achieve an optimal tradeoff between channel exploration and channel exploitation (the process to utilize observed channel opportunities). In the literature, the channel sensing and access process has been modeled as a multi-armed bandit problem (MABP) [12].

2.3.2.1 Multi-Armed Bandit Problem

The name “multi-arm bandit problem" comes from an imaginary slot machine equipped which multiple hands (also called *arms*). For each arm, if it is pulled, a reward, which is a random variable, can be obtained. The distribution of the random variable is unknown. And each arm has a unique distribution of the random variable reward. Time is partitioned into slots. At each time slot, a player can pull an arm and obtain a reward. The player can also estimate the statistical behaviors of the arms based on the observations over time. The decision (i.e., which arm to pull) at each time slot is made based on the player's observations in previous time slots and the player's estimate of the arms' statistical information. The player is to maximize a long-term accumulated reward.

Such an MABP is a classical stochastic adaptive control problem. In the problem, the player faces a dilemma between limited *information* obtained and effective *control* expected. Here the information means the distribution of the random variable reward of each arm, which needs be learned by pulling the arm for a sufficiently long time. The control refers to the goal of maximization of the long-term accumulated reward, which means that it is preferred to select the arm with the largest reward to pull. As a simple example, consider a slot machine with two arms: Arm A and Arm B. The probability density function of the random variable rewards of the two arms are denoted as p_A and p_B, respectively. At each time slot, say Slot j, the player selects to pull either Arm A or Arm B, and obtains a reward A_j or B_j (which are random variables with probability density function p_A and p_B, respectively). Denoting the reward at Slot j as x_j, the MABP is to decide on the arm selection at each slot such that the expected value of the sum $S_t \triangleq \sum_{j=1}^{t} x_j$ is maximized when t is sufficiently large.

2.3.2.2 Modeling of OCSA by MABP

The OCSA with multiple channels can be modeled by an MABP as follows. Consider that a secondary user has multiple potential channels. At each time slot, the secondary user can sense one channel during the sensing period, and access the channel in the transmission period of the slot if the channel is sensed free. Therefore, the secondary user is the player, the channels are the arms to be pulled, the amount of data transmitted over the transmission period is the reward, and the distribution of the state (idle/busy) of each channel is the distribution of the reward of each arm. The objective is to maximize the amount of transmitted secondary traffic.

For an MABP, the regret $R(t)$ until time instant t is the expected difference between the actual reward of an arm-selection rule and the reward of a genie-aided rule that has known statistical information of the arms [87]. It is proved in [88] that, for any *adaptive allocation rule*[1], the regret is at least $\mu \ln t$ when $t \to \infty$, where the factor μ is determined by the statistical information of arms. A rule that achieves the lower bound of μ is called *efficiently optimal*, and a rule with regret $O(\ln t)$ is called *order optimal*. For OCSA in CRNs, reference [12] derives order optimal rules to well coordinate the balance between channel exploration and exploitation, with the assumption of perfect channel sensing. Although not efficiently optimal, the rules are *sample mean based index rules* [89], and their implementation is much simpler than the efficiently optimal rules given in [88]. Moreover, a regret bound is also observed with finite t [2] in rules in [12], while no such bound is observed for finite t in the efficiently optimal rules in [88]. A distributed cognitive sensing problem is investigated and formulated as an *adversary bandit problem* in [29], where no statistical assumption is made on channel states. Multi-user OCSA in a distributed manner is investigated in [31], modeled as an MABP with multiple players. In the above existing research efforts for OCSA in CRNs, perfect channel sensing is assumed, and each secondary user can utilize all observed spectrum opportunities (i.e., can access all sensed-free channels).

[1] This means the decisions of the rule are only based on observations in the history [88].

[2] In this report, when we say “finite t", it means sufficiently large and finite t.

Chapter 3

CRNs with Statistical Information of Primary Traffic and with Imperfect Spectrum Sensing: The Optimal Sets of Channels to Sense and to Access

In this chapter, OCSA problem of a secondary user with multiple potential primary channels is investigated. The secondary user can sense a limited number of channels, and channel sensing is imperfect. If the secondary user can access all channels sensed free, it is proved that the secondary user should sense the channels with the largest rewards, where the reward of a channel is the reward that the secondary user can acquire if it senses the channel and accesses the channel when the channel is sensed free. If the secondary user can access only a limited number of sensed-free channels, in general it may not be optimal to sense the channels with the largest rewards. However and interestingly, for some special cases (for example, when all the channels have the same detection probability), simple rules are given for the optimal selection of channels to sense. For the general case, methods are given to reduce the searching complexity for the optimal set of channels to sense.[1]

[1] *A version of this chapter has been published in IEEE Wireless Communications Letters, 1: 133-136 (2012).*

3.1 Introduction

In a CRN, secondary users need to first detect possible primary activities, usually by spectrum sensing, and then access the spectrum if no primary activities are detected. When there are multiple potential primary spectrum bands (called *primary channels*) for a secondary user, the secondary user needs to decide which channel(s) to sense and access, and how they are sensed and accessed. In [7], at the beginning of a time slot, a secondary user sequentially senses the channels one after another, until a free channel or a free channel with good channel quality is found. Then the secondary user transmits in the channel within the remaining duration of the slot. The optimal order for sensing the channels is derived. In [90], sensing order when the channel gain information is known is studied. In [91], sensing order is jointly designed with sensing strategy (to specify when to stop sensing and start secondary transmission) and power allocation, to maximize energy efficiency. In [12], a secondary user senses one channel or senses multiple channels simultaneously at the beginning of a time slot, and accesses sensed-free channel(s) in the remaining duration of the slot. The channel sensing and access problem is formulated as a multi-armed bandit problem. Sensing time optimization is investigated in [8] for a single-channel case and in [11] for a multiple-channel case. Aggregated opportunistic throughput is maximized in [15].

In the above existing works, it is assumed that the secondary user can access all sensed-free channels. In this research, we consider a system when a secondary user simultaneously senses a limited number of channels (e.g., by wideband spectrum sensing technique discussed in [92]) at the beginning of a time slot and uses the remaining duration in the slot for data transmission. Different from existing works, spectrum sensing is imperfect, and the secondary user can only access up to a limited number of sensed-free channels in a slot.[2] We aim at deciding which channels to sense so that the secondary user can gain the maximal reward. We find that, when the secondary user can access all sensed-free channels, the secondary user should sense the channels that have the largest rewards (the definition of reward of a channel is to be given in Section 3.2). However and interestingly, if the secondary user can only access up to a limited number of sensed-free channels at a time, it may not be optimal to sense the channels with the largest rewards, and thus, exhaustive search may

[2] For example, in a voice conversation, the secondary user may only have limited packets to send during a time period. As another example, as shown in [93], due to energy constraint, the secondary user may not be able to access all sensed-free channels.

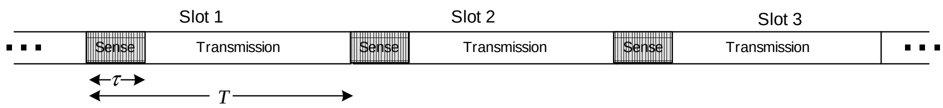

Fig. 3.1. The slotted time structure.

be needed to find the optimal set of channels to sense. Some simple rules are given for the optimal selection of channel set to sense in some special cases. And a property is also given for the general case, which helps to simplify the search for the optimal channel set to sense.

3.2 System Model

Consider a secondary user with N potential primary channels, denoted as Channel 1, Channel 2, ..., Channel N. Similar to [7], [8], [12], [90], time is partitioned into slots, each with fixed duration T. Each slot is further divided into a sensing period with duration τ and a data transmission period with duration $(T-\tau)$, as shown in Fig. 3.1. The secondary user can sense M $(\leq N)$ channels simultaneously in the sensing period, and subsequently in the data transmission period it can access up to K $(\leq M)$ channels that are sensed free. To protect primary users, the secondary user is not permitted to access channels sensed busy. Since sensing is imperfect, for sensing of Channel i $(i = 1, 2, ..., N)$, let P_d^i denote the detection probability (i.e., probability of detecting primary activities that do exist), and P_f^i denote the false alarm probability (i.e., the probability of mistakenly estimating presence of primary activities that actually do not exist).

At each slot, say Slot j, Channel i $(i = 1, 2, ..., N)$ is free with probability θ_i. Let $S_i(j) = 1$ and $S_i(j) = 0$ denote that Channel i is free and busy, respectively; and if Channel i is sensed, let $X_i(j) = 1$ and $X_i(j) = 0$ denote that Channel i is sensed to be free and busy, respectively. The probability of Channel i being sensed free is denoted $f(\theta_i) = \theta_i(1-P_f^i)+(1-\theta_i)(1-P_d^i) = \theta_i(P_d^i-P_f^i)+1-P_d^i$. The i.i.d. model of busy/idle states of a channel over time slots is considered, i.e., for each channel, the channel state varies independently across time slots. The N primary channels have independent channel states.

In a slot, if the secondary user accesses a channel that is sensed free, it can transmit B bits in the data transmission period. Define *reward* as the *successfully* transmitted bits in a slot. So if there is a missed detection of primary activities in an accessed channel, then the reward is 0. For a channel, say Channel i, define its reward as the expected reward the secondary user can acquire if the secondary user senses Channel i and accesses it when it is sensed free. Also define *conditional reward* of Channel i as the expected reward to access Channel i conditioned on that Channel i is sensed free. So for Channel i, its reward is given as $B\theta_i(1-P_f^i)$, and its conditional reward is given as $B\mathbb{E}[S_i(j)|X_i(j)=1]=\frac{B\theta_i(1-P_f^i)}{f(\theta_i)}$, where $\mathbb{E}[\cdot]$ denotes expectation.

Since the secondary user does not sense all the channels, and may not access all channels sensed free, the secondary user has two decisions: which channels to sense, and which sensed-free channels to access. For the second decision, it is apparent that: if the number of channels sensed free is not more than K, then all channels sensed free are accessed; otherwise, the secondary user should access the K channels with the K largest conditional rewards. Therefore, in this research, we focus on the first decision of the secondary user: which M channels to sense. Our objective is to maximize the expected reward of the secondary user in a slot (say Slot j), given as:

$$\max_{\mathcal{M}\subseteq\mathcal{N}} R_{\mathcal{M}} \triangleq \mathbb{E}\Big[B \max_{\mathcal{K}\subseteq\mathcal{I}_{\mathcal{M}}} \sum_{i\in\mathcal{K}} \mathbb{E}\left[S_i(j)|X_i(j)=1\right]\Big] \tag{3.1}$$

where $\mathcal{N}=\{1,2,...,N\}$, $\mathcal{M}$ denotes the set of channels to sense, $\mathcal{I}_{\mathcal{M}}$ is the set of channels that are sensed free if channels in $\mathcal{M}$ are sensed, $\mathcal{K}$ denotes the set of channels to access, $R_{\mathcal{M}}$ denotes the *reward of* $\mathcal{M}$, defined as expected reward of the secondary user if it senses the channels in set $\mathcal{M}$ and accesses up to K sensed-free channels with the largest conditional rewards. In (3.1), the outer expectation is for $\mathcal{I}_{\mathcal{M}}$, while the inner expectation is for $S_i(j)$, $i\in\mathcal{K}$. For (3.1) and subsequent equations, we have $|\mathcal{M}|=M$, and $|\mathcal{K}|\leq K$, where $|\cdot|$ means the cardinality of a set.

3.3 Optimal Selection of Channels to Sense

We consider two cases: full channel access with $K=M$ (i.e., the secondary user accesses all channels that are sensed free), and partial channel access with $K<M$.

3.3.1 Full Channel Access ($K = M$)

Full channel access also means $\mathcal{K} = \mathcal{I}_{\mathcal{M}}$. Then we have the following theorem for the optimal set of channels to sense.

Theorem 3.1. The optimal set of channels to sense, denoted as $\mathcal{M}^*$, consists of M channels with the M largest values of $\theta_i(1 - P_f^i)$, $i \in \mathcal{N}$.

Proof. Since $\mathcal{K} = \mathcal{I}_{\mathcal{M}}$, problem in (3.1) is equivalent to

$$\max_{\mathcal{M}\subseteq\mathcal{N}} \mathbb{E}\Big[B \sum_{i\in\mathcal{I}_{\mathcal{M}}} \mathbb{E}\left[S_i(j)|X_i(j)=1\right]\Big] = \max_{\mathcal{M}\subseteq\mathcal{N}} B \sum_{i\in\mathcal{M}} \mathbb{P}(i \in \mathcal{I}_{\mathcal{M}})\mathbb{E}\left[S_i(j)|X_i(j)=1\right]$$

$$= \max_{\mathcal{M}\subseteq\mathcal{N}} B \sum_{i\in\mathcal{M}} f(\theta_i) \cdot \frac{\theta_i(1-P_f^i)}{f(\theta_i)} = \max_{\mathcal{M}\subseteq\mathcal{N}} B \sum_{i\in\mathcal{M}} \theta_i(1-P_f^i)$$

where $\mathbb{P}(\cdot)$ means probability of an event.

Therefore, to maximize the expected reward of the secondary user, the secondary user should sense the M channels with the M largest values of $\theta_i(1 - P_f^i)$, $i \in \mathcal{N}$. $\square$

Since $B\theta_i(1 - P_f^i)$ is reward of Channel i, Theorem 3.1 is intuitive: to sense the M channels with the M largest rewards.

3.3.2 Partial Channel Access ($K < M$)

For partial channel access, our first question is: does an intuitive rule as that in Theorem 3.1 exist? Unfortunately, for partial channel access, it may not be optimal to sense the M channels with the M largest rewards. Here is an example. Let $N = 5$, $M = 4$, $K = 1$, and $B = 1$. $(\theta_1, \theta_2, ..., \theta_5) = (0.83, 0.47, 0.34, 0.39, 0.51)$, $(P_d^1, P_d^2, ..., P_d^5) = (0.7, 0.6, 0.55, 0.65, 0.9)$, $(P_f^1, P_f^2, ..., P_f^5) = (0.4, 0.2, 0.15, 0.3, 0.5)$. By choosing the $M = 4$ channels with largest $\theta_i(1-P_f^i)$, we have a set $\mathcal{M} = \{1, 2, 3, 4\}$, and the secondary user's expected reward in a slot is 0.746. However, by exhaustive search, the optimal set of channels to sense is $\mathcal{M} = \{1, 2, 3, 5\}$, with which the secondary user's expected reward in a slot is 0.768. Therefore, although the intuitive rule (i.e., selecting the channels with largest rewards) is optimal for full channel access, it may not be optimal for partial channel access. The reason is: in partial channel access, the secondary user may not access a channel that is sensed free.

Since the intuitive rule is not optimal in general for partial channel access, it seems that exhaustive search may be needed to find the optimal set of channels to sense. However, interestingly, in some special cases, some simple rules exist, as shown in Sections 3.3.2.1 and 3.3.2.2, while in the general case, the searching complexity for the optimal channel set to sense can be reduced according to a property, as shown in Section 3.3.2.3.

3.3.2.1 With Homogeneous Sensing

Here homogeneous sensing means all the channels have the same detection probability (i.e., $P_d^i = P_d,\ i \in \mathcal{N}$) and the same false alarm probability (i.e., $P_f^i = P_f,\ i \in \mathcal{N}$). Without loss of generality, we assume $\theta_1 > \theta_2 > ... > \theta_N$ in Section 3.3.2.1. We have the following theorem.

Theorem 3.2. With homogeneous sensing, the optimal set of channels to sense is $\{1, 2, ..., M\}$.

Proof. We use proof by contradiction. Assume that the optimal set of channels to sense, $\mathcal{M}^*$, is not $\{1, 2, ..., M\}$. Denote $\mathcal{M}^*$ as $\mathcal{M}^* = \{n_1, n_2, ..., n_M\}$ with $n_1 < n_2 < ... < n_M$. It means $\theta_{n_1} > \theta_{n_2} > ... > \theta_{n_M}$. Note that, with homogeneous sensing, if a channel has a larger free probability θ_i, it also has a larger conditional reward.

Since $\mathcal{M}^*$ is not $\{1, 2, .., M\}$, there exists a channel, denoted Channel $l \in \{1, 2, ..., M\}$, such that $l \notin \mathcal{M}^*$. Then l is smaller than at least one element in $\mathcal{M}^*$, and thus, there exists $k \in \{1, 2, ..., M\}$ such that $n_{k-1} < l < n_k$.[3] It also means $\theta_{n_{k-1}} > \theta_l > \theta_{n_k}$.

Now we derive an expression for $R_{\mathcal{M}^*}$, the reward of $\mathcal{M}^*$. Consider sensing of the $(M-1)$ channels in $\mathcal{M}^* \backslash \{n_k\}$. Define the set of channels sensed free, $\mathcal{I}_{\mathcal{M}^* \backslash \{n_k\}}$, as the *sensing result*, and denote the set of all 2^{M-1} possible sensing result realizations as $\mathbb{U}$. Further, we have $\mathbb{U} = \mathbb{U}_1 \cup \mathbb{U}_2$, where $\mathbb{U}_1$ is the set of sensing result realizations in which the number of sensed-free channels among Channels $n_1, n_2, ..., n_{k-1}$ is less than K, and $\mathbb{U}_2$ is the set of sensing result realizations in which the number of sensed-free channels among

[3] If $k = 1$, then we have $l < n_1$, which can be treated similarly.

Channels $n_1, n_2, ..., n_{k-1}$ is equal to or more than K. Then the reward of $\mathcal{M}^*$ is given as

$$
\begin{aligned}
&R_{\mathcal{M}^*} = \\
&f(\theta_{n_k})\Big(\sum_{\mathcal{U}\in\mathbb{U}_1}\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})\Big(\tfrac{B\theta_{n_k}(1-P_f)}{f(\theta_{n_k})}+r_{\mathcal{U}}^{K-1}\Big)+\sum_{\mathcal{U}\in\mathbb{U}_2}\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})r_{\mathcal{U}}^{K}\Big)\\
&+\Big(1-f(\theta_{n_k})\Big)\Big(\sum_{\mathcal{U}\in\mathbb{U}_1}\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})r_{\mathcal{U}}^{K}+\sum_{\mathcal{U}\in\mathbb{U}_2}\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})r_{\mathcal{U}}^{K}\Big)\\
&=\sum_{\mathcal{U}\in\mathbb{U}_1}\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})\Big(\theta_{n_k}\big((1-P_f)B-(P_d-P_f)(r_{\mathcal{U}}^{K}-r_{\mathcal{U}}^{K-1})\big)-(1-P_d)(r_{\mathcal{U}}^{K}-r_{\mathcal{U}}^{K-1})\Big)\\
&\quad+\sum_{\mathcal{U}\in\mathbb{U}}\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})r_{\mathcal{U}}^{K}
\end{aligned}
$$

where $\mathcal{U}$ is a sensing result realization of sensing channels in $\mathcal{M}^*\backslash\{n_k\}$, $r_{\mathcal{U}}^{K-1}$ and $r_{\mathcal{U}}^{K}$ are the rewards by accessing up to $(K-1)$ and K channels in $\mathcal{U}$, respectively, that have the largest conditional rewards.

In $\mathcal{M}^*$, if we replace Channel n_k with Channel l, we get set $\{n_1, n_2, ..., n_{k-1}, l, n_{k+1}, ..., n_M\}$. Similarly, its reward is given as

$$
\begin{aligned}
R_{\{n_1,n_2,...,n_{k-1},l,n_{k+1},...,n_M\}} &= \sum_{\mathcal{U}\in\mathbb{U}_1}\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})\Big(\theta_l\big((1-P_f)B\\
&-(P_d-P_f)(r_{\mathcal{U}}^{K}-r_{\mathcal{U}}^{K-1})\big)-(1-P_d)(r_{\mathcal{U}}^{K}-r_{\mathcal{U}}^{K-1})\Big)+\sum_{\mathcal{U}\in\mathbb{U}}\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})r_{\mathcal{U}}^{K}.
\end{aligned}
$$

Then the difference of the rewards of $\mathcal{M}^*$ and $\{n_1, n_2, ..., n_{k-1}, l, n_{k+1}, ..., n_M\}$ is given as

$$
\begin{aligned}
&R_{\{n_1,n_2,...,n_{k-1},l,n_{k+1},...,n_M\}}-R_{\mathcal{M}^*}=\\
&(\theta_l-\theta_{n_k})(1-P_f)B\sum_{\mathcal{U}\in\mathbb{U}_1}\Big[\mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}}=\mathcal{U})\cdot\Big(1-\frac{P_d-P_f}{(1-P_f)B}(r_{\mathcal{U}}^{K}-r_{\mathcal{U}}^{K-1})\Big)\Big]>0
\end{aligned}
\tag{3.2}
$$

where the inequality comes from $\theta_l > \theta_{n_k}$ and the following fact. According to the definition of $r_{\mathcal{U}}^{K}$ and $r_{\mathcal{U}}^{K-1}$, their difference is no more than the conditional reward of Channel 1 (which has the largest conditional reward), which means:

$$
r_{\mathcal{U}}^{K}-r_{\mathcal{U}}^{K-1}\le B\mathbb{E}[S_1(j)|X_1(j)=1]=\tfrac{B\theta_1(1-P_f)}{\theta_1(P_d-P_f)+1-P_d}<\tfrac{B(1-P_f)}{P_d-P_f}. \tag{3.3}
$$

Inequality (3.2) contradicts the assumption that $\mathcal{M}^*$ is optimal. This completes the proof. □

3.3.2.2 With Common Detection Probability

Now we consider a special case when all the channels have a common detection probability (i.e., $P_d^i = P_d, i \in \mathcal{N}$)[4] but have different false alarm probabilities. Without loss of generality, we assume $\theta_1 > \theta_2 > ... > \theta_N$ in Section 3.3.2.2. We have the following theorem.

Theorem 3.3. When all the channels have a common detection probability, if both $\theta_1 > \theta_2 > ... > \theta_N$ and $\theta_1(1-P_f^1) > \theta_2(1-P_f^2) > ... > \theta_N(1-P_f^N)$ are satisfied, the optimal set of channels to sense is $\{1, 2, ..., M\}$.

Proof. From $\theta_1 > \theta_2 > ... > \theta_N$ and $\theta_1(1-P_f^1) > \theta_2(1-P_f^2) > ... > \theta_N(1-P_f^N)$, we have the following for the conditional rewards of the channels

$$\mathbb{E}[S_1(j)|X_1(j) = 1] > \mathbb{E}[S_2(j)|X_2(j) = 1] > ... > \mathbb{E}[S_N(j)|X_N(j) = 1]. \quad (3.4)$$

Then the optimality of $\{1, 2, ..., M\}$ can be proved as follows.

We use proof by contradiction. Assume that the optimal set of channels to sense, $\mathcal{M}^* = \{n_1, n_2, ..., n_M\}$ with $n_1 < n_2 < ... < n_M$, is not $\{1, 2, .., M\}$. Then there exists a channel, denoted Channel $l \in \{1, 2, ..., M\}$, such that $l \notin \mathcal{M}^*$. And there exists $k \in \{1, 2, ..., M\}$ such that $n_{k-1} < l < n_k$.

Similar to (3.3.2.1), the reward of $\mathcal{M}^*$ is given as

$$\begin{aligned} R_{\mathcal{M}^*} = &\sum_{\mathcal{U}\in\mathbb{U}_1} \mathbb{P}(\mathcal{I}_{\mathcal{M}^*\setminus\{n_k\}} = \mathcal{U})\Big(\theta_{n_k}(1-P_f^{n_k})\big(B-(r_{\mathcal{U}}^K - r_{\mathcal{U}}^{K-1})\big) \\ &+ \theta_{n_k}(1-P_d)(r_{\mathcal{U}}^K - r_{\mathcal{U}}^{K-1}) - (1-P_d)(r_{\mathcal{U}}^K - r_{\mathcal{U}}^{K-1})\Big) + \sum_{\mathcal{U}\in\mathbb{U}} \mathbb{P}(\mathcal{I}_{\mathcal{M}^*\setminus\{n_k\}} = \mathcal{U}) r_{\mathcal{U}}^K \end{aligned} \quad (3.5)$$

where $\mathcal{U}$, $\mathbb{U}$, $\mathbb{U}_1$, $r_{\mathcal{U}}^{K-1}$, and $r_{\mathcal{U}}^K$ have the same definitions as in proof of Theorem 3.2.

In $\mathcal{M}^*$, if we replace Channel n_k with Channel l, we get set $\{n_1, n_2, ..., n_{k-1}, l, n_{k+1}, ..., n_M\}$. Similar to (3.5), its reward is given as

$$R_{\{n_1,n_2,...,n_{k-1},l,n_{k+1},...,n_M\}} = \sum_{\mathcal{U}\in\mathbb{U}_1} \mathbb{P}(\mathcal{I}_{\mathcal{M}^*\setminus\{n_k\}} = \mathcal{U})\Big(\theta_l(1-P_f^l)\big(B-(r_{\mathcal{U}}^K - r_{\mathcal{U}}^{K-1})\big)$$

[4] As an example, if it is required that the detection probability in each channel is above a common threshold so as to protect primary users, then the secondary user may set its detection probability in each channel as the common threshold value.

$$+ \theta_l(1-P_d)(r_{\mathcal{U}}^K - r_{\mathcal{U}}^{K-1}) - (1-P_d)(r_{\mathcal{U}}^K - r_{\mathcal{U}}^{K-1})\Big)$$
$$+ \sum_{\mathcal{U}\in\mathbb{U}} \mathbb{P}(\mathcal{I}_{\mathcal{M}^*\backslash\{n_k\}} = \mathcal{U}) r_{\mathcal{U}}^K. \tag{3.6}$$

Since $\theta_l > \theta_{n_k}$, $\theta_l(1-P_f^l) > \theta_{n_k}(1-P_f^{n_k})$, $0 \le r_{\mathcal{U}}^K - r_{\mathcal{U}}^{K-1} \le B$, it can be seen that $R_{\{n_1,n_2,\ldots,n_{k-1},l,n_{k+1},\ldots,n_M\}} > R_{\mathcal{M}^*}$, which contradicts the assumption that $\mathcal{M}^*$ is optimal. □

Theorem 3.2 indicates that for homogeneous sensing, the secondary user should sense the M channels with the M largest free probabilities. Theorem 3.3 indicates that, for a case when only detection probabilities are common while false alarm probabilities are different, if adding factor $(1-P_f^i)$ does not affect the ordering of the free probabilities of the channels, then the secondary user should still sense the M channels with the M largest free probabilities.

Next, for determining of the optimal $\mathcal{M}$, we have a definition of *preferred channel* as follows. Channel a is said to be preferred to Channel b if the following condition is satisfied: if Channel b is in the optimal $\mathcal{M}$, then Channel a should be also in the optimal $\mathcal{M}$.

Theorem 3.4. When all the channels have a common detection probability, for any pair of channels, a channel is preferred to the other channel if it has both larger free probability (i.e., θ_i) and larger reward (i.e., $B\theta_i(1-P_f^i)$) than those of the other channel, respectively.

Proof. We use proof by contradiction. For Channels i_1 and i_2, assume $\theta_{i_1} > \theta_{i_2}$ and $\theta_{i_1}(1-P_f^{i_1}) > \theta_{i_2}(1-P_f^{i_2})$. Denote $\mathcal{M}^*$ as the optimal set of channels to sense, and $i_1 \notin \mathcal{M}^*$, $i_2 \in \mathcal{M}^*$. Denote the conditional rewards of Channels i_1 and i_2 as

$$y_1 = \frac{B\theta_{i_1}(1-P_f^{i_1})}{\theta_{i_1}(1-P_f^{i_1}) + (1-\theta_{i_1})(1-P_d)}$$

and

$$y_2 = \frac{B\theta_{i_2}(1-P_f^{i_2})}{\theta_{i_2}(1-P_f^{i_2}) + (1-\theta_{i_2})(1-P_d)}$$

respectively. Then $y_1 > y_2$.

In $\mathcal{M}^*$, if we replace Channel i_2 by Channel i_1, we get set $\mathcal{M}^\dagger$.

For sensing of the $(M-1)$ channels in $\mathcal{M}^*\backslash\{i_2\}$, denote $\mathcal{U}$ as a sensing result realization (i.e., the set of sensed-free channels). We partition $\mathcal{U}$ into three subsets: $\mathcal{U}_1$ includes the sensed-free channels whose conditional rewards are larger than $y_1, \mathcal{U}_2$ includes the sensed-free channels whose conditional rewards are less than or equal to y_1 and larger than y_2, and $\mathcal{U}_3$ includes the sensed-free channels whose conditional rewards are less than or equal to y_2.

When the sensing result of the $(M-1)$ channels in $\mathcal{M}^*\backslash\{i_2\}$ is fixed as $\mathcal{U}$, denote the reward of $\mathcal{M}^*$ and $\mathcal{M}^\dagger$ as $R^*_{\mathcal{U}}$ and $R^\dagger_{\mathcal{U}}$, respectively. Next, we derive expressions of $R^*_{\mathcal{U}}$ and $R^\dagger_{\mathcal{U}}$. Let $r^{K-1}_{\mathcal{U}}$ and $r^K_{\mathcal{U}}$ be the rewards by accessing up to $(K-1)$ channels and K channels in $\mathcal{U}$, respectively, that have the largest conditional rewards. We have the following three possible scenarios for $\mathcal{U}$.

- Scenario with $|\mathcal{U}_1| \geq K$: We have $R^*_{\mathcal{U}} = R^\dagger_{\mathcal{U}} = r^K_{\mathcal{U}}$.
- Scenario with $|\mathcal{U}_1| < K$ and $|\mathcal{U}_1| + |\mathcal{U}_2| \geq K$: We have $R^*_{\mathcal{U}} = r^K_{\mathcal{U}}$, $R^\dagger_{\mathcal{U}} = f(\theta_{i_1})(r^{K-1}_{\mathcal{U}} + y_1) + (1 - f(\theta_{i_1}))r^K_{\mathcal{U}}$. Then we have $R^\dagger_{\mathcal{U}} - R^*_{\mathcal{U}} = f(\theta_{i_1})\big(y_1 - (r^K_{\mathcal{U}} - r^{K-1}_{\mathcal{U}})\big) \geq 0$.
- Scenario with $|\mathcal{U}_1| + |\mathcal{U}_2| < K$: We have

$$R^*_{\mathcal{U}} = f(\theta_{i_2})(r^{K-1}_{\mathcal{U}} + y_2) + (1 - f(\theta_{i_2}))r^K_{\mathcal{U}} = f(\theta_{i_2})\big(y_2 - (r^K_{\mathcal{U}} - r^{K-1}_{\mathcal{U}})\big) + r^K_{\mathcal{U}}, \tag{3.7}$$

$$R^\dagger_{\mathcal{U}} = f(\theta_{i_1})(r^{K-1}_{\mathcal{U}} + y_1) + (1 - f(\theta_{i_1}))r^K_{\mathcal{U}} = f(\theta_{i_1})\big(y_1 - (r^K_{\mathcal{U}} - r^{K-1}_{\mathcal{U}})\big) + r^K_{\mathcal{U}}. \tag{3.8}$$

  If $f(\theta_{i_1}) \geq f(\theta_{i_2})$, then we have $R^\dagger_{\mathcal{U}} > R^*_{\mathcal{U}}$ since $y_1 > y_2$, $y_1 - (r^K_{\mathcal{U}} - r^{K-1}_{\mathcal{U}}) > 0$, and $y_2 - (r^K_{\mathcal{U}} - r^{K-1}_{\mathcal{U}}) \geq 0$. If $f(\theta_{i_1}) < f(\theta_{i_2})$, then (3.7) and (3.8) can be rewritten as

$$R^*_{\mathcal{U}} = B\theta_{i_2}(1 - P^{i_2}_f) - f(\theta_{i_2})(r^K_{\mathcal{U}} - r^{K-1}_{\mathcal{U}}) + r^K_{\mathcal{U}}$$

  and

$$R^\dagger_{\mathcal{U}} = B\theta_{i_1}(1 - P^{i_1}_f) - f(\theta_{i_1})(r^K_{\mathcal{U}} - r^{K-1}_{\mathcal{U}}) + r^K_{\mathcal{U}}$$

  respectively, which also lead to $R^\dagger_{\mathcal{U}} > R^*_{\mathcal{U}}$ since $\theta_{i_1}(1 - P^{i_1}_f) > \theta_{i_2}(1 - P^{i_2}_f)$ and $r^K_{\mathcal{U}} - r^{K-1}_{\mathcal{U}} \geq 0$.

The probability of the last scenario is nonzero. Therefore, averaged on all possible $\mathcal{U}$'s, the reward of $\mathcal{M}^\dagger$ is larger than the reward of $\mathcal{M}^*$, which contradicts the assumption that $\mathcal{M}^*$

is the optimal set of channels to sense. This completes the proof. □

Based on Theorem 3.4, the following corollary can be proved straightforwardly.

Corollary 3.1. For N channels with common detection probability and with $\theta_1 > \theta_2 > ... > \theta_N$, if there exists $k \in \{1, 2, ..., M\}$ such that $\theta_1(1 - P_f^1) > \theta_2(1 - P_f^2) > ... > \theta_k(1 - P_f^k) > \max\{\theta_{k+1}(1 - P_f^{k+1}), \theta_{k+2}(1 - P_f^{k+2}), ..., \theta_N(1 - P_f^N)\}$, then $\{1, 2, ..., k\}$ is a subset of the optimal $\mathcal{M}$.

3.3.2.3 Property for the General Case

For the general case with neither common detection probability nor common false alarm probability, we have the following theorem, whose proof is similar to that of Theorem 3.4, and is omitted.

Theorem 3.5. For any pair of channels, a channel is preferred to the other channel if it has both larger sensed-free probability $f(\theta_i)$ and larger conditional reward than those of the other channel, respectively.

Theorem 3.5 can be used to reduce the searching complexity for the optimal $\mathcal{M}$ in the general case. Based on Theorem 3.5, the following corollaries can be straightforwardly proved.

Corollary 3.2. If $(n_1, n_2, ..., n_N)$ is a permutation of $(1, 2, ..., N)$, and if conditions $f(\theta_{n_1}) > f(\theta_{n_2}) > ... > f(\theta_{n_N})$ and $\frac{\theta_{n_1}(1-P_f^{n_1})}{f(\theta_{n_1})} > \frac{\theta_{n_2}(1-P_f^{n_2})}{f(\theta_{n_2})} > ... > \frac{\theta_{n_N}(1-P_f^{n_N})}{f(\theta_{n_N})}$ are satisfied, the optimal set of channels to sense, denoted as $\mathcal{M}^*$, is $\{n_1, n_2, ..., n_M\}$.

Corollary 3.3. If $(n_1, n_2, ..., n_N)$ is a permutation of $(1, 2, ..., N)$, with $f(\theta_{n_1}) > f(\theta_{n_2}) > ... > f(\theta_{n_N})$, and if there exists $k \in \{1, 2, ..., M\}$ such that $\frac{\theta_{n_1}(1-P_f^{n_1})}{f(\theta_{n_1})} > \frac{\theta_{n_2}(1-P_f^{n_2})}{f(\theta_{n_2})} > ... > \frac{\theta_{n_k}(1-P_f^{n_k})}{f(\theta_{n_k})} > \max\left\{\frac{\theta_{n_{k+1}}(1-P_f^{n_{k+1}})}{f(\theta_{n_{k+1}})}, \frac{\theta_{n_{k+2}}(1-P_f^{n_{k+2}})}{f(\theta_{n_{k+2}})}, ..., \frac{\theta_{n_N}(1-P_f^{n_N})}{f(\theta_{n_N})}\right\}$, then $\{n_1, n_2, ..., n_k\}$ is a subset of the optimal $\mathcal{M}$.

3.4 Performance Evaluation

Next we show numerical results to demonstrate the impact of the selection of channels to sense. Consider 4 channels with channel free probabilities $(\theta_1, \theta_2, \theta_3, \theta_4) = (0.650, 0.727, 0.852, 0.918)$. Three cases are investigated: homogeneous case with $P_d = 0.7$ and $P_f =$

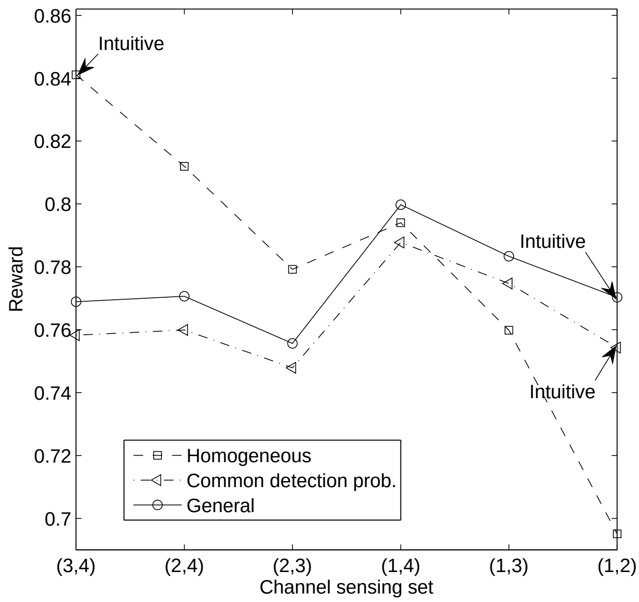

Fig. 3.2. Reward of the set of channels to sense.

0.3; common detection probability case with $P_d = 0.7$ and $(P_f^1, P_f^2, P_f^3, P_f^4) = (0.1, 0.28, 0.39, 0.43)$; and general case with $(P_d^1, P_d^2, P_d^3, P_d^4) = (0.8, 0.8, 0.8, 0.95)$ and $(P_f^1, P_f^2, P_f^3, P_f^4) = (0.1, 0.28, 0.39, 0.43)$. The secondary user can sense two channels and access one channel. Fig. 3.2 shows the reward of different set of channels to sense in the three cases. The reward of the intuitive rule (i.e., the two channels with the two largest $\theta_i(1 - P_f^i)$ are sensed) is also indicated. It can be seen that the intuitive rule is optimal in the homogeneous case, and not optimal in the other two cases.

3.5 Conclusion

In this chapter, we have found some interesting results for the optimal set of channels to be sensed by a secondary user. When the secondary user can utilize all sensed-free channels, the intuitive rule is optimal. However, this intuitive rule is not optimal in general when the secondary user can only access up to a limited number of sensed-free channels. Interestingly, we have found some simple rules for the optimal set of channels to sense in some special cases. And for the general case, we have provided a guideline to reduce the searching complexity for the optimal channel set to sense.

In this chapter, the secondary user can transmit B bits at each slot over each free channel. It is interesting to consider that the number of transmitted bits at a slot over different free channels are different. For full channel access, it can be found optimal to sense the M channels with largest rewards, defined as $B_i\theta_i(1 - P_f^i)$, in which B_i is the number of transmitted bits at a time slot on Channel i if it is free. But the case for partial channel access is more complicated, and deserves further investigation.

Chapter 4

Channel Exploration and Exploitation with Imperfect Spectrum Sensing in CRNs: i.i.d. Model of Channel Busy/Idle States over Time

In this chapter, the problem of OCSA in CRNs when the sensing is imperfect and a secondary user can access up to a limited number of channels at a time is investigated. Primary users' statistical information is assumed to be unknown, and therefore, a secondary user needs to learn the information online during channel sensing and access process, which means learning loss, also referred to as regret, is inevitable. For each channel, the busy/idle state is independent from one slot to another. In this research, the case when all potential channels can be sensed simultaneously is investigated first. The channel access process is modeled as an MABP with side observation. And channel access rules are derived and theoretically proved to have asymptotically finite regret. Then the case when the secondary user can sense only a limited number of channels at a time is investigated. The channel sensing and access process is modeled as a bi-level MABP. It is shown that any adaptive rule has at least logarithmic regret. Then we derive OCSA rules and theoretically prove that they have logarithmic regret asymptotically and with finite time. The effectiveness of the derived rules is validated by computer simulation.[1]

4.1 Introduction

In a spectrum-overlay CRN, OCSA is used, in which the secondary users search for spectrum holes through sensing, and utilize the observed spectrum opportunities for their data transmission. Optimal OCSA when the secondary users have statistical information of primary users, such as information of idle probabilities of primary channels, has been addressed in [8], [11], [13]–[15] and Chapter 3 of this thesis, to maximize transmission capacity, optimize transmission power efficiency, etc. However, research on the optimal OCSA without *a priori* statistical knowledge of primary channels is still in its infancy. The research challenge is how to achieve the optimal tradeoff between channel exploration (the process to sense the channels so as to learn the statistical information) and channel exploitation (the process to utilize observed channel opportunities). If statistical information of primary channels is known in advance, a secondary user can select the optimal channel(s) to sense and subsequently access sensed-idle channel(s). However, without such information, a learning process is needed, and the secondary user should also explore suboptimal channels through sensing to learn statistical information of those channels. Therefore, learning loss is expected, compared to the case that the secondary user always selects the optimal channel(s).

As discussed in Chapter 2, in the literature, the OCSA process in CRNs has been modeled as MABPs [12], [29]–[31], [87]–[89], which assume that channel sensing is perfect and a secondary user can access all sensed-idle channels. Different from these existing research efforts, this chapter investigates OCSA when (i) imperfect channel sensing is assumed and (ii) a secondary user can access up to a limited number of channels simultaneously (i.e., it may not use all observed spectrum opportunities at a time). Our motivation for (i) is that channel sensing is always imperfect in a real network. And our motivation for (ii) is that the secondary user may have energy and/or hardware constraints, and thus, the number of channels that can be accessed is limited. A similar setting is also adopted in [16], [17], [93], [94].[2]

---

[1] *The version of this chapter was published in IEEE Journal on Selected Areas in Communications, 31: 429-441 (2013).*

[2] Actually the case when a secondary user can access an unlimited number of channels can be viewed as a special case of our work.

Therefore, unlike existing OCSA research where there is only one decision (i.e., to decide which channels to sense, and subsequently access all sensed-idle channels), we have two decisions in the OCSA in our work: to decide which channels to sense; and if a number of channels are sensed idle, to decide which channels to access. Two cases are considered in our work:

- Case I: when a secondary user can sense all potential channels simultaneously[3], referred to as *full channel sensing*;
- Case II: when a secondary user can sense a limited number of potential channels simultaneously, referred to as *partial channel sensing*.

4.2 Case I: with Full Channel Sensing

Consider the same slotted time structure as in Fig. 3.1, where time is partitioned into slots, and the duration of each slot is T. For a secondary user, there are N potential primary channels, denoted as Channels $1, 2, ..., N$, respectively. In each slot, Channel i ($i \in \{1, 2, ..., N\}$) is idle (i.e., without primary activities) with probability θ_i, and θ_i is unknown by the secondary user. Let $S_i(j) = 1$ and $S_i(j) = 0$ denote Channel i is idle and busy, respectively, at Slot j. The i.i.d. model of busy/idle states of a channel over time slots is considered. In other words, for each channel, the channel state (busy or idle) varies independently from a slot to another. And the N channels have independent channel states. Similar settings are adopted in [7], [8], [11], [12], [93].

Each slot consists of a sensing period with duration τ and data transmission period with duration $T - \tau$. For each slot, during the sensing period the secondary user senses all the N channels. Among all the sensed-idle channels, the secondary user can access (i.e., transmit its data over) up to K channels in the data transmission period. For each accessed channel, reward (i.e., the information bits the secondary user can transmit in a slot) is normalized to 1.

During the sensing in Slot j, denote $\mathbf{X}(j) = (X_1(j), X_2(j), ..., X_N(j))$ as the *sensing observation* of the N channels, where $X_i(j) = 1$ and $X_i(j) = 0$ mean that Channel i is sensed to be idle and busy, respectively. Since sensing errors are inevitable, similar to Chapter 3, we let P_d^i denote the detection probability of Channel i (i.e., the probability

[3]This may be achieved by the wideband spectrum sensing technique discussed in [92].

of detecting the primary user activity if there is primary user activity), and P_f^i denote the false-alarm probability of Channel i (i.e., the probability of mistakenly estimating that the primary user is active when there is actually no primary user activity). For Channel i at Slot j, the probability that it is sensed idle (i.e., $X_i(j) = 1$) is given as $f(\theta_i) \triangleq (1-P_f^i)\theta_i + (1-P_d^i)(1-\theta_i)$. So, conditioned on that Channel i is sensed idle at Slot j, the *conditional reward* of Channel i if it is accessed at Slot j can be calculated as $\mathbb{E}[S_i(j)|X_i(j) = 1] = \frac{(1-P_f^i)\theta_i}{f(\theta_i)}$, where $\mathbb{E}[\cdot]$ denotes expectation.

Since the secondary user senses all the N channels, the only decision of the secondary user to make is on which channel(s) to access based on its sensing observation. To protect primary users, only channels sensed idle can be accessed. Since primary users' statistical information $\boldsymbol{\Theta} \triangleq (\theta_1, \theta_2, ..., \theta_N)$ is unknown, online learning is needed for the secondary user to estimate $\boldsymbol{\Theta}$. In the following, we first investigate the situation of single channel access (i.e., $K = 1$, the secondary user can access only one channel at a slot), and subsequently extend the research result to the situation of multiple channel access (i.e., $K \geq 2$, the secondary user can access more than one channel simultaneously at a slot).

4.2.1 Single Channel Access at a Slot ($K = 1$)

To evaluate the performance of a channel access rule, we use the performance of a genie-aided rule (in which the channel statistical information $\boldsymbol{\Theta}$ is known) as a benchmark for comparison. For the genie-aided rule, let $\mathcal{I}(j)$ denote the set of channels sensed idle at Slot j. Then among channels in $\mathcal{I}(j)$, the secondary user should access the channel with the maximal expected reward. If Channel i is sensed idle (i.e., $i \in \mathcal{I}(j)$), Channel i's expected reward is actually the conditional reward $\mathbb{E}[S_i(j)|X_i(j) = 1]$. Therefore, the secondary user should access the sensed-idle channel with the maximal conditional reward, i.e., $\max\limits_{i \in \mathcal{I}(j)} \mathbb{E}\left[S_i(j)|X_i(j) = 1\right]$. And until Slot t, the expected reward of the genie-aided rule is thus given as $\sum\limits_{j=1}^{t} \mathbb{E}\Big[\max\limits_{i \in \mathcal{I}(j)} \mathbb{E}\left[S_i(j)|X_i(j) = 1\right]\Big]$, where the outer expectation is for $\mathcal{I}(j)$ (totally there are 2^N different possible sets of sensed-idle channels), and the inner expectation is to calculate the conditional reward of Channel i.

For any adaptive allocation rule denoted ψ, where $\psi(j) = i$ means Channel i is decided to be accessed at Slot j, the expected reward until Slot t is $\mathbb{E}\big[\sum\limits_{j=1}^{t}\sum\limits_{i=1}^{N} \mathbb{E}[S_i(j)|X_i(j) = 1] \times \mathbb{I}[\psi(j) = i]\big]$, where the outer expectation is for $\psi(j)$, and $\mathbb{I}[\cdot]$ means the indicator function.

The regret (also the learning loss) of rule ψ until Slot t, defined as the difference between the expected rewards of ψ and the genie-aided rule, is given as

$$R(t,\psi) = \sum_{j=1}^{t}\mathbb{E}\big[\max_{i\in\mathcal{I}(j)}\mathbb{E}[S_i(j)|X_i(j)=1]\big]-\mathbb{E}\Big[\sum_{j=1}^{t}\sum_{i=1}^{N}\mathbb{E}[S_i(j)|X_i(j)=1]\mathbb{I}[\psi(j)=i]\Big]. \tag{4.1}$$

Since the secondary user can sense all the channels before selecting a channel to access, the channel access process can be modeled as an *MABP with side observation* [95]. For an MABP, it is extremely hard to derive an optimal channel access strategy such that the regret is minimized. Therefore, researchers instead focus on regret bound in asymptotic sense. For example, in [12], asymptotically order optimal rules are derived such that the regret is $O(\ln t)$ when $t \to \infty$. In our research, we also focus on channel access rules with good asymptotic performance such as asymptotically finite regret. Note that for *two-armed* bandit problem with side observation, reference [95] gives a rule with asymptotically finite regret under *direct information* setting. In our work, we derive a rule with asymptotically finite regret for our *multi-armed* bandit problem with side observation, as follows.

Define *sensing result* as the set of channels that are sensed idle. For sensing of the N channels, we have 2^N possible sensing results. At Slot t, we keep a 2^N-sized vector of the sample-mean probabilities of the 2^N sensing results, denoted $\mathbf{P}_s = (P_{s,1}, P_{s,2}, ..., P_{s,2^N})$, in which $P_{s,i}$ is the sample-mean probability of the ith sensing result, given as (the number of slots until Slot t in which the ith sensing result happens)$/t$. We also denote $\mathbf{P}_{\mathbf{\Theta}'} = (P_{\mathbf{\Theta}',1}, P_{\mathbf{\Theta}',2}, ..., P_{\mathbf{\Theta}',2^N})$ as a 2^N-sized vector of the probabilities of the 2^N sensing results by assuming that $\mathbf{\Theta}' = (\theta'_1, \theta'_2, ..., \theta'_N)$ is the vector of the channel idle probabilities. For example, if the 2nd sensing result is the set of Channels 1 and 2, then $P_{\mathbf{\Theta}',2} = f(\theta'_1)f(\theta'_2)\prod_{i=3}^{2N}(1-f(\theta'_i))$, where $f(\theta_i)$ is defined at the beginning of Section 4.2.

Our proposed channel access rule is shown in Algorithm 4.1.

In Line 7 of Algorithm 4.1, $\mathbb{E}_{\hat{\mathbf{\Theta}}}[S_i(t)|X_i(t)=1]$ means the conditional reward of Channel i, by assuming that $\hat{\mathbf{\Theta}}$ is the vector of the channel idle probabilities. In other words, $\mathbb{E}_{\hat{\mathbf{\Theta}}}[S_i(t)|X_i(t)=1] = \frac{(1-P_f^i)\hat{\theta}_i}{f(\hat{\theta}_i)}$, where $\hat{\theta}_i$ is the ith element in $\hat{\mathbf{\Theta}}$. Here we select the channel with the largest conditional reward by using $\hat{\mathbf{\Theta}}$ as the estimation of $\mathbf{\Theta}$.

Remark 4.1. In Algorithm 4.1, the key concept is the 2^N-sized vectors of the probabilities

Algorithm 4.1 Single Channel Access with Full Channel Sensing at Slot t

1: Sense N channels, record sensing observation $\mathbf{X}(t) = (X_1(t), X_2(t), ..., X_N(t))$, and update $\mathbf{P}_s$.

2: Construct candidate set $\mathcal{C}(t)$ of the form

$$\mathcal{C}(t) = \left\{ \mathbf{\Theta}^\dagger : ||\mathbf{P}_{\mathbf{\Theta}^\dagger} - \mathbf{P}_s||_2 \leq \inf_{\mathbf{\Theta}' \in (0,1]^N} ||\mathbf{P}_{\mathbf{\Theta}'} - \mathbf{P}_s||_2 + \frac{1}{t} \right\}$$

where $|| \cdot ||_2$ is the L_2-norm of a vector, given as $||\mathbf{x}||_2 = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$ for a vector $\mathbf{x} = (x_1, x_2, ..., x_n)$.

3: Arbitrarily pick up $\hat{\mathbf{\Theta}} \in \mathcal{C}(t)$, as the estimation of $\mathbf{\Theta}$.

4: **if** the set of channels sensed idle at Slot t, $\mathcal{I}(t)$, is empty **then**

5: Do not access any channel at Slot t.

6: **else**

7: Access Channel $i^* = \arg\max_{i \in \mathcal{I}(t)} \mathbb{E}_{\hat{\mathbf{\Theta}}} \left[S_i(t) | X_i(t) = 1 \right]$.

of the 2^N sensing results. At each slot, $\mathbf{P}_s$ is the sample-mean vector until the slot. And by assuming any channel idle probability vector $\mathbf{\Theta}'$, we can calculate $\mathbf{P}_{\mathbf{\Theta}'}$. Here we use the L_2-norm of $(\mathbf{P}_{\mathbf{\Theta}'} - \mathbf{P}_s)$ to represent the distance of the two vectors $\mathbf{P}_{\mathbf{\Theta}'}$ and $\mathbf{P}_s$. The first term on the right hand side of the inequality in Line 2 of the algorithm is the infinimum of the distance from $\mathbf{P}_{\mathbf{\Theta}'}$ to the sample-mean vector $\mathbf{P}_s$. In the algorithm, any vector $\mathbf{\Theta}^\dagger$ can be taken as the estimated channel idle probability vector, as long as the distance of its associated vector $\mathbf{P}_{\mathbf{\Theta}^\dagger}$ to the sample-mean vector $\mathbf{P}_s$ is "not far" from the infinimum distance. Here "not far" means that the difference of the two distances is bounded by $1/t$. When t is small, the sample-mean vector $\mathbf{P}_s$ may not be accurate to reflect the vector of real probabilities of the 2^N sensing results, and thus, we search $\mathbf{P}_{\mathbf{\Theta}^\dagger}$ in a large region around $\mathbf{P}_s$ (i.e., $1/t$ is large). When t becomes large, the sample-mean vector $\mathbf{P}_s$ becomes accurate enough to reflect the vector of real probabilities of the 2^N sensing results, and thus, we search $\mathbf{P}_{\mathbf{\Theta}^\dagger}$ in a small region around $\mathbf{P}_s$ (i.e., $1/t$ is small).

Theorem 4.1. Algorithm 4.1 achieves asymptotically finite regret; that is, $\limsup_{t \to \infty} R(t) < \infty$.

Proof. Recall that $\mathbf{\Theta}$ is the vector of real channel idle probabilities, and in Line 3 of Algorithm 4.1, $\hat{\mathbf{\Theta}}$ is used to estimate $\mathbf{\Theta}$. Denote the sensing result at Slot t as $u(t)$, where $u(t) = l$ ($1 \leq l \leq 2^N$) means the lth sensing result. Denote $k_{\mathbf{\Theta}}(u(t))$ and $k_{\hat{\mathbf{\Theta}}}(u(t))$ as the best channel which has the largest conditional reward (conditioned on that the sensing result is $u(t)$) when $\mathbf{\Theta}$ and $\hat{\mathbf{\Theta}}$ are used as the vector of channel idle probabilities, respectively.

By following Algorithm 4.1, the probability of wrong access (i.e. access a suboptimal channel) is

$$\mathbb{P}\left(k_{\hat{\boldsymbol{\Theta}}}(u(t)) \neq k_{\boldsymbol{\Theta}}(u(t))\right) \leq \mathbb{P}\left(\exists u \in \mathcal{U},\ k_{\hat{\boldsymbol{\Theta}}}(u) \neq k_{\boldsymbol{\Theta}}(u)\right) \tag{4.2}$$

where $\mathbb{P}(\cdot)$ means probability, and $\mathcal{U} = \{1, 2, 3..., 2^N\}$.

Define a set $\mathcal{C}_e \triangleq \{\boldsymbol{\Theta}' : \exists u \in \mathcal{U},\ k_{\boldsymbol{\Theta}'}(u) \neq k_{\boldsymbol{\Theta}}(u)\}$. Then (4.2) is equivalent to

$$\mathbb{P}\left(k_{\hat{\boldsymbol{\Theta}}}(u(t)) \neq k_{\boldsymbol{\Theta}}(u(t))\right) \leq \mathbb{P}\left(\hat{\boldsymbol{\Theta}} \in \mathcal{C}_e\right). \tag{4.3}$$

Define $\varepsilon \stackrel{\triangle}{=} \inf\limits_{\boldsymbol{\Theta}' \in \mathcal{C}_e} ||\mathbf{P}_{\boldsymbol{\Theta}'} - \mathbf{P}_{\boldsymbol{\Theta}}||_2$. Then we have $\varepsilon > 0$ with proof given in Appendix 4.A.

We first consider that an event $\left\{||\mathbf{P}_{\boldsymbol{\Theta}} - \mathbf{P}_s||_2 < \frac{\varepsilon}{3}\right\}$ happens. From Algorithm 4.1, we have

$$||\mathbf{P}_{\hat{\boldsymbol{\Theta}}} - \mathbf{P}_s||_2 \leq \inf_{\boldsymbol{\Theta}' \in (0,1]^N} ||\mathbf{P}_{\boldsymbol{\Theta}'} - \mathbf{P}_s||_2 + \frac{1}{t} \leq ||\mathbf{P}_{\boldsymbol{\Theta}} - \mathbf{P}_s||_2 + \frac{1}{t} < \frac{\varepsilon}{3} + \frac{1}{t}. \tag{4.4}$$

When t is large enough such that $\frac{1}{t} \leq \frac{\varepsilon}{3}$, from (4.4) we have

$$||\mathbf{P}_{\boldsymbol{\Theta}} - \mathbf{P}_{\hat{\boldsymbol{\Theta}}}||_2 \leq ||\mathbf{P}_{\boldsymbol{\Theta}} - \mathbf{P}_s||_2 + ||\mathbf{P}_{\hat{\boldsymbol{\Theta}}} - \mathbf{P}_s||_2 < \varepsilon \tag{4.5}$$

which means $\hat{\boldsymbol{\Theta}} \notin \mathcal{C}_e$ from the definition of ε. It also means that, if $\hat{\boldsymbol{\Theta}} \in \mathcal{C}_e$, then we should have $||\mathbf{P}_{\boldsymbol{\Theta}} - \mathbf{P}_s||_2 \geq \frac{\varepsilon}{3}$. Then we have

$$\mathbb{P}\left(\hat{\boldsymbol{\Theta}} \in \mathcal{C}_e\right) \leq \mathbb{P}\left(||\mathbf{P}_{\boldsymbol{\Theta}} - \mathbf{P}_s||_2 \geq \frac{\varepsilon}{3}\right) \leq a(t) \stackrel{\triangle}{=} (t+1)^{2^N} e^{-t \inf\limits_{\mathbf{P}'_s \in \mathcal{B}} \sum\limits_{i \in \mathcal{U}} P'_{s,i} \ln \frac{P'_{s,i}}{P_{\boldsymbol{\Theta},i}}} \tag{4.6}$$

where the second inequality comes from the Sanov Theorem (i.e., Theorem 2.1.10) in [96], and $\mathcal{B}$ denotes a vector space $\left\{\mathbf{P}'_s : ||\mathbf{P}_{\boldsymbol{\Theta}} - \mathbf{P}'_s||_2 \geq \frac{\varepsilon}{3}\right\}$, which is closed, and $\mathbf{P}'_s = (P'_{s,1}, P'_{s,2}, ..., P'_{s,2^N})$.

For the exponent in the expression of $a(t)$, we have

$$\sum_{i \in \mathcal{U}} P'_{s,i} \ln \frac{P'_{s,i}}{P_{\boldsymbol{\Theta},i}} = \sum_{i \in \mathcal{U}} P_{\boldsymbol{\Theta},i} \frac{P'_{s,i}}{P_{\boldsymbol{\Theta},i}} \ln \frac{P'_{s,i}}{P_{\boldsymbol{\Theta},i}} \geq \left(\sum_{i \in \mathcal{U}} P_{\boldsymbol{\Theta},i} \frac{P'_{s,i}}{P_{\boldsymbol{\Theta},i}}\right) \ln \sum_{i \in \mathcal{U}} P_{\boldsymbol{\Theta},i} \frac{P'_{s,i}}{P_{\boldsymbol{\Theta},i}} = 0 \tag{4.7}$$

where the inequality comes from the Jensen's inequality and the fact that $x \ln x$ is a convex function. In addition, $\sum\limits_{i\in\mathcal{U}} P'_{s,i} \ln \frac{P'_{s,i}}{P_{\mathbf{\Theta},i}}$ is continuous and strictly convex, which, together with $\varepsilon > 0$ and (4.7), leads to $\inf\limits_{\mathbf{P}'_s\in\mathcal{B}} \sum\limits_{i\in\mathcal{U}} P'_{s,i} \ln \frac{P'_{s,i}}{P_{\mathbf{\Theta},i}} > 0$. And thus, from the definition of $a(t)$ given in (4.6), we have $\lim\limits_{t\to\infty} \frac{a(t+1)}{a(t)} < 1$.

From (4.3) and (4.6), we have $\mathbb{P}\left(k_{\hat{\mathbf{\Theta}}}\left(u(t)\right) \neq k_{\mathbf{\Theta}}(u(t))\right) \leq a(t)$ when $\frac{1}{t} \leq \frac{\varepsilon}{3}$. So for regret $R(t)$ of Algorithm 4.1, we have

$$\begin{aligned}
&\limsup_{t\to\infty} R(t) \\
&\leq c_0 \sum_{j=1}^{\lfloor\frac{3}{\varepsilon}\rfloor} \mathbb{P}\left(k_{\hat{\mathbf{\Theta}}}\left(u(j)\right) \neq k_{\mathbf{\Theta}}(u(j))\right) + c_0 \lim_{t\to\infty} \sum_{j=\lfloor\frac{3}{\varepsilon}\rfloor+1}^{t} \mathbb{P}\left(k_{\hat{\mathbf{\Theta}}}\left(u(j)\right) \neq k_{\mathbf{\Theta}}(u(j))\right) \\
&\leq c_0 \left\lfloor \frac{3}{\varepsilon} \right\rfloor + c_0 \lim_{t\to\infty} \sum_{j=\lfloor\frac{3}{\varepsilon}\rfloor+1}^{t} a(j) < \infty
\end{aligned} \tag{4.8}$$

where $\lfloor\cdot\rfloor$ is a floor function, c_0 denotes the largest possible reward loss due to wrong access in a slot, which is finite, and the last inequality comes from $\lim\limits_{t\to\infty} \frac{a(t+1)}{a(t)} < 1$ by the ratio test of convergency of a series.

Therefore, by following Algorithm 4.1, asymptotically finite regret is achieved. □

Theorem 4.1 indicates the performance of Algorithm 4.1 is surprisingly good through full channel sensing prior to channel access. As a comparison, in the rules derived in [12] where the secondary user senses one channel with perfect sensing, performance of $R(t) \sim O(\ln t)$ is achieved, which means the regret goes to infinity when $t \to \infty$.

Algorithm 4.1 suffers from high complexity in the construction of candidate set $\mathcal{C}(t)$ in each slot, especially when t is large. To reduce complexity, an alternative channel access rule with linear complexity is introduced, as given in Algorithm 4.2.

Algorithm 4.2 Single Channel Access with Full Channel Sensing at Slot t

1: Sense N channels, and get sensing observation $\mathbf{X}(t)$.
2: Estimate the idle probability of Channel i ($i \in \{1, 2, ..., N\}$) to be $\hat{\theta}_i(t) = \frac{\frac{1}{t}\sum_{j=1}^{t} X_i(j) + P_d^i - 1}{P_d^i - P_f^i}$.
3: **if** the set of channels sensed idle at Slot t, $\mathcal{I}(t)$, is empty **then**
4: Do not access any channel at Slot t.
5: **else**
6: Access Channel $i^* = \arg\max\limits_{i\in\mathcal{I}(t)} \mathbb{E}_{\hat{\mathbf{\Theta}}}\left[S_i(t)|X_i(t) = 1\right]$ where $\hat{\mathbf{\Theta}} = (\hat{\theta}_1, \hat{\theta}_1, ..., \hat{\theta}_N)$.

Remark 4.2. For Channel i with idle probability θ_i, detection probability P_d^i, and false alarm probability P_f^i, the expectation of the sample-mean $\frac{1}{t}\sum_{j=1}^{t} X_i(j)$ is $f(\theta_i) = (P_d^i - P_f^i)\theta_i + 1 - P_d^i$. Thus, we use unbiased estimator of θ_i, given as $\hat{\theta}_i(t) = \frac{\frac{1}{t}\sum_{j=1}^{t} X_i(j) + P_d^i - 1}{P_d^i - P_f^i}$. Note that, in a CRN, normally a large detection probability (e.g., not less than 0.9 in the IEEE 802.22 proposal) and a small false alarm probability (e.g., not more than 0.1 in the IEEE 802.22 proposal) are required. Therefore, we have $P_d^i \neq P_f^i$.

Theorem 4.2. Algorithm 4.2 achieves asymptotically finite regret.

Proof. Define function $g_i(x) \triangleq \frac{(1-P_f^i)x}{(1-P_f^i)x+(1-P_d^i)(1-x)}$, $0 < x < 1$. Then the conditional reward of Channel i at Slot j, i.e., $\mathbb{E}[S_i(j)|X_i(j) = 1] = \frac{(1-P_f^i)\theta_i}{(1-P_f^i)\theta_i+(1-P_d^i)(1-\theta_i)}$, can be expressed as $g_i(\theta_i)$. Without loss of generality, assume $g_1(\theta_1) > g_2(\theta_2) > ... > g_N(\theta_N)$. Then the genie-aided rule is to access the channel with the smallest index among all sensed-free channels. Define $\varepsilon \triangleq \min_{1\leq i<N} \big(g_i(\theta_i) - g_{i+1}(\theta_{i+1})\big)$, which is the minimal difference of the conditional rewards of any two channels.

At Slot t, $g_i(\hat{\theta}_i(t))$ is the conditional reward of Channel i based on the estimated channel idle probability $\hat{\theta}_i(t)$ in Algorithm 4.2. It is apparent that, if for each channel $i \in \{1, 2, ..., N\}$ we have $|g_i(\hat{\theta}_i(t)) - g_i(\theta_i)| < \frac{\varepsilon}{2}$, then Algorithm 4.2 will make correct access (i.e., access the same channel as the genie-aided rule does) at Slot t. This means

$$\mathbb{P}(\text{correct access at Slot } t) \geq \prod_{i=1}^{N} \mathbb{P}\left(|g_i(\hat{\theta}_i(t)) - g_i(\theta_i)| < \frac{\varepsilon}{2}\right). \tag{4.9}$$

Note that $g_i(x)$ is a continuous and strictly increasing function of $x \in (0, 1)$. Denote its inverse function as $g_i^{-1}(\cdot)$, which is a continuous and strictly increasing function within $(g_i(0), g_i(1))$. Denote $\nu_i \triangleq \min\big(g_i^{-1}(g_i(\theta_i) + \frac{\varepsilon}{2}) - \theta_i,\ \theta_i - g_i^{-1}(g_i(\theta_i) - \frac{\varepsilon}{2})\big)$,[4] which is a positive value. So, if $|\hat{\theta}_i(t) - \theta_i| < \nu_i$, then we have $|g_i(\hat{\theta}_i(t)) - g_i(\theta_i)| < \frac{\varepsilon}{2}$. This means

$$\mathbb{P}\left(|g_i(\hat{\theta}_i(t)) - g_i(\theta_i)| < \frac{\varepsilon}{2}\right) \geq \mathbb{P}\left(|\hat{\theta}_i(t) - \theta_i| < \nu_i\right). \tag{4.10}$$

[4]Note that the range of $g_i(x)$ is $(g_i(0), g_i(1))$. So when $g_i(\theta_i) + \frac{\varepsilon}{2} > g_i(1)$, we set $g_i^{-1}(g_i(\theta_i) + \frac{\varepsilon}{2}) = 1$; when $g_i(\theta_i) - \frac{\varepsilon}{2} < g_i(0)$, we set $g_i^{-1}(g_i(\theta_i) - \frac{\varepsilon}{2}) = 0$.

Based on Chernoff-Hoeffding bound, we have

$$\mathbb{P}\left(|\hat{\theta}_i(t) - \theta_i| < \nu_i\right) \geq 1 - 2e^{-2\nu_i^2 t}. \tag{4.11}$$

Based on (4.9)-(4.11), we have

$$\begin{aligned} &\mathbb{P}(\text{wrong access at Slot } t) = 1 - \mathbb{P}(\text{correct access at Slot } t) \\ &\leq 1 - \prod_{i=1}^{N}\left(1 - 2e^{-2\nu_i^2 t}\right) \leq 1 - (1 - 2e^{-2\nu^2 t})^N \\ &= 2Ne^{-2\nu^2 t} - \sum_{l=2}^{N}\binom{N}{l}\left(-2e^{-2\nu^2 t}\right)^l \end{aligned} \tag{4.12}$$

where $\nu = \min_{1\leq i\leq N} \nu_i$.

Then for regret $R(t)$ of Algorithm 4.2, we have

$$\begin{aligned} &\limsup_{t\to\infty} R(t) \leq \limsup_{t\to\infty} \sum_{j=1}^{t} c_0 \mathbb{P}\left(\text{wrong access at Slot } t\right) \\ &\leq \limsup_{t\to\infty} \sum_{j=1}^{t} c_0 \left(2Ne^{-2\nu^2 t} - \sum_{l=2}^{N}\binom{N}{l}\left(-2e^{-2\nu^2 t}\right)^l\right) \\ &< \infty \end{aligned}$$

where c_0 denotes the largest possible reward loss by accessing a wrong channel at a slot. □

4.2.2 Multiple Channel Access at a Slot ($K > 1$)

Assume the secondary user can simultaneously access up to $K(> 1)$ channels at a slot. Therefore, if the number of channels sensed idle at a slot is less than or equal to K, then all those sensed-idle channels are accessed by the secondary user; otherwise, K channels are selected among the sensed-idle channels to be accessed by the secondary user.

We still use the performance of a genie-aided rule with $\mathbf{\Theta}$ known as a benchmark for comparison. Until Slot t, the expected reward of the genie-aided rule is given as

$$\sum_{j=1}^{t} \mathbb{E}\Big[\max_{\mathcal{K}(j)\subset\mathcal{I}(j), |\mathcal{K}(j)|\leq K} \sum_{i\in\mathcal{K}(j)} \mathbb{E}[S_i(j)|X_i(j) = 1]\Big]$$

where $\mathcal{I}(j)$ denotes the set of channels sensed idle at Slot j and $\mathcal{K}(j)$ denotes the set of channels to be accessed at Slot j. The outer expectation is for $\mathcal{I}(j)$, while the inner expectation is to calculate the conditional reward for Channel i.

For any adaptive allocation rule Ψ with multiple channel access, where $\Psi(j)$ denotes the set of channels to be accessed at Slot j, the expected reward until Slot t is $\mathbb{E}\big[\sum_{j=1}^{t}\sum_{i=1}^{N}\mathbb{E}[S_i(j)|X_i(j)=1]\mathbb{I}[i\in\Psi(j)]\big]$.

The regret of rule Ψ is given as

$$R(t,\Psi) = \sum_{j=1}^{t}\mathbb{E}\Big[\max_{\mathcal{K}(j)\subset\mathcal{I}(j),|\mathcal{K}(j)|\le K}\sum_{i\in\mathcal{K}(j)}\mathbb{E}[S_i(j)|X_i(j)=1]\Big]$$
$$-\,\mathbb{E}\big[\sum_{j=1}^{t}\sum_{i=1}^{N}\mathbb{E}[S_i(j)|X_i(j)=1]\mathbb{I}[i\in\Psi(j)]\big].$$

For multiple channel access, we modify Line 7 in Algorithm 4.1 and Line 6 in Algorithm 4.2 as follows: if $|\mathcal{I}(t)|\le K$, then access all channels in $\mathcal{I}(t)$; otherwise, among all the channels in $\mathcal{I}(t)$, access the K channels with the largest K values of $\mathbb{E}_{\hat{\Theta}}[S_i(t)|X_i(t)=1]$. It can be proved that the resulted algorithms have asymptotically finite regret, by using similar proofs to those of Theorems 4.1 and 4.2.

4.3 Case II: with Partial Channel Sensing

Still consider N channels. At a slot, the secondary user can sense $M(<N)$ channels and can access up to $K(\le M)$ channels among the sensed-idle channels. Therefore, we have a *bi-level MABP*: the first level is to decide which M channels to sense; and the second level is to decide, among the sensed-idle channels, which up to K channels to access. The arms played in the two levels are different, which makes the problem much more challenging than classical MABP. To the best of our knowledge, a general bi-level MABP is still an open problem. In the following, we provide solutions to our particular bi-level MABP. Possible extension of our solutions to a more general bi-level MABP is to be investigated in our future work.

Unlike Case I where we have common channel access rules for homogeneous sensing (i.e., $P_d^i = P_d$, $P_f^i = P_f$, $\forall i\in\{1,2,...,N\}$) and heterogeneous sensing (i.e., for each channel, say Channel i, we have distinct setting $\{P_d^i, P_f^i\}$), the homogeneous sensing and heterogeneous sensing need to be treated in different ways in Case II, as discussed in Section 4.3.1 and 4.3.2, respectively.

4.3.1 Homogeneous Sensing

Consider $P_d^i = P_d$, $P_f^i = P_f$, $\forall i \in \{1, 2, ..., N\}$. Without loss of generality, we assume $\theta_1 > \theta_2 > ... > \theta_N$.

We still use the performance of a genie-aided rule as a benchmark for comparison. From Theorem 3.2, it can be seen that the genie-aided rule should always sense $\mathcal{M}^* = \{1, 2, ..., M\}$. So until Slot t, the expected reward of the genie-aided rule is given as

$$U^*(t) = \sum_{j=1}^{t} \mathbb{E}\left[\max_{\mathcal{K}(j)\subset\mathcal{I}_{\mathcal{M}^*}(j), |\mathcal{K}(j)|\leq K} \sum_{i\in\mathcal{K}(j)} \mathbb{E}\left[S_i(j)|X_i(j) = 1\right]\right]$$

where $\mathcal{I}_{\mathcal{M}^*}(j)$ denotes the set of sensed-idle channels at Slot j if the channels in $\mathcal{M}^*$ are sensed, and $\mathcal{K}(j)$ denotes the set of channels to access at Slot j.

In the following, we investigate single channel access ($K = 1$) and multiple channel access ($K > 1$), respectively.

4.3.1.1 Single Channel Access at a Slot ($K = 1$)

The expected reward of the genie-aided rule until Slot t is:

$$U^*(t) = \sum_{j=1}^{t} \mathbb{E}\left[\max_{i\in\mathcal{I}_{\mathcal{M}^*}(j)} \mathbb{E}\left[S_i(j)|X_i(j) = 1\right]\right]. \tag{4.13}$$

Compared with the genie-aided rule, regret of a single channel access rule ϕ, in which $\phi(j)$ denotes the channel to be accessed at Slot j, is given as

$$R(t, \phi) = U^*(t) - \mathbb{E}\left[\sum_{j=1}^{t}\sum_{i=1}^{N} \mathbb{E}[S_i(j)|X_i(j) = 1]\mathbb{I}[\phi(j) = i]\right]. \tag{4.14}$$

Unlike Case I in Section 4.2, we cannot expect asymptotically finite regret $R(t)$. The reason is as follows. For partial channel sensing, consider a *perfect scenario* in which all sensed-idle channels are to be accessed and all sensings are perfect. It is shown in Theorem 3.1 in [88] and Lemma 2 in [12] that the perfect scenario has a lower bound of $O(\ln t)$ on $R(t)$ as $t \to \infty$. Compared with such perfect scenario, our considered system suffers extra learning loss due to the possibility that among the sensed-idle channels, not those with the largest conditional rewards are accessed. Therefore, the regret of any rule in our Case II has a lower bound of $O(\ln t)$.

Note that references [88] and [89] give rules with regret $O(\ln t)$ when $t \to \infty$. However, performance of the rules with finite t is still unclear. In the following, using the UCB1 (here UCB stands for Upper Confidence Bound) in [97], we derive an OCSA rule that has regret $R(t) \sim O(\ln t)$ with $t \to \infty$ and with finite t. Note that the original UCB1 cannot be directly applied to our research problem, because, if it is directly applied, there is only one decision, i.e., which channels to sense at a slot. Since in our research problem there are two decisions (which channels to sense, and which channel to access among the sensed-idle channels), we have nontrivial extensions to the original UCB1.

At each slot (say Slot t), the secondary user keeps records $\mathbf{T}(t) = (T_1(t), T_2(t), ..., T_N(t))$ and $\mathbf{Y}(t) = (Y_1(t), Y_2(t), ..., Y_N(t))$, where $T_i(t)$ is the number of slots in which Channel i has been sensed until Slot t, and Y_i is the number of slots in which Channel i has been sensed idle until Slot t. The proposed OCSA rule is given in Algorithm 4.3.

Algorithm 4.3 Single Channel Access with Homogeneous Sensing in Case II (Partial Channel Sensing)

1: Sense all N channels by using $\lceil \frac{N}{M} \rceil$ slots (where $\lceil \cdot \rceil$ is a ceiling function). At each slot, randomly select one sensed-idle channel to access. Update $\mathbf{T}$ and $\mathbf{Y}$ at each slot.
2: **for** each subsequent Slot t **do**
3: Estimate θ_i ($i = 1, 2, ..., N$) by $\hat{\theta}_i(t) = \frac{\frac{Y_i(t-1)}{T_i(t-1)} + P_d - 1}{P_d - P_f}$, and determine channel set $\mathcal{M}(t)$ to sense, which includes channels with the largest M indices $\hat{\theta}_i(t) + \frac{1}{P_d - P_f}\sqrt{\frac{2\ln(t-1)}{T_i(t-1)}}$.
4: Sense channels in $\mathcal{M}(t)$. Let $\mathcal{I}(t)$ denote the set of sensed-idle channels. Update $\mathbf{T}(t)$ and $\mathbf{Y}(t)$.
5: **if** $\mathcal{I}(t)$ is nonempty **then**
6: Access Channel $i^* = \arg\max_{i \in \mathcal{I}(t)} \left\{ \hat{\theta}_i(t) + \frac{1}{P_d - P_f}\sqrt{\frac{2\ln(t-1)}{T_i(t-1)}} \right\}$.
7: **else**
8: Do not access any channel at Slot t.

Remark 4.3. Similar to Algorithm 4.2, we use an unbiased estimation of θ_i. When determining the channels to sense, we add the extra term $\frac{1}{P_d - P_f}\sqrt{\frac{2\ln(t-1)}{T_i(t-1)}}$ in the index. Its purpose is to guarantee that each channel is sensed in a sufficient number of slots, as follows. When a channel, say Channel i, is not sufficiently sensed, this channel has a relatively small $T_i(t-1)$, while all channels have the same $\ln(t-1)$. So Channel i is very likely to have a larger index than other channels, which gives Channel i a larger chance to be selected to sense. Consider two channels, Channel i_1 and Channel i_2, as an example. For Channel i_1, its $T_{i_1}(t)$ does not grow in the scale of $O(\ln t)$, and thus, the extra term (which is very likely larger than 1) dominates in the index of Channel i_1. For Channel i_2, its $T_{i_2}(t)$ grows in the scale of $O(\ln t)$, and thus, the term $\hat{\theta}_{i_2}(t)$ dominates in the index of Channel i_2. Then it is very likely that Channel i_1 has a larger index than Channel i_2, which gives Channel i_1 more chance to be selected. Therefore, the extra term in the index implies that the time to sense each channel is at least $O(\ln t)$.[5] When all the channels' $T_i(t)$'s ($i = 1, 2, ...N$) grow at least in the scale of $O(\ln t)$, the terms $\hat{\theta}_i(t)$'s tend to dominate in the indices, and the secondary user tends to select the channel with the largest $\hat{\theta}_i(t)$.

And when the sensing of a channel is not accurate enough (small P_d or large P_f), the extra term will give the channel a larger chance to be sensed. This is desired because more sensings are needed to estimate θ_i if the sensing is not accurate enough.

<u>Theorem</u> 4.3. The regret $R(t)$ of Algorithm 4.3 is $O(\ln t)$ with $t \to \infty$ and with finite t.

Proof. Recall that we assume $\theta_1 > \theta_2 > ... > \theta_N$, and for the genie-aided rule, $\mathcal{M}^* = \{1, 2, ..., M\}$ is the optimal set of channels to sense. Then for any rule, the expected reward loss in a slot (say Slot j) is bounded by the maximal expected reward of the genie-aided rule in the slot, given as $\Delta \triangleq \mathbb{E}\big[\max_{i\in\mathcal{M}^*} \frac{\theta_i(1-P_f)}{f(\theta_i)} X_i(j)\big]$, where $f(\theta_i) = (1 - P_f^i)\theta_i + (1 - P_d^i)(1 - \theta_i)$ is the probability that Channel i is sensed idle.

Recall that in Algorithm 4.3, $\mathcal{M}(j)$ denotes the set of channels to sense at Slot j. So until Slot t, the regret $R(t)$ of Algorithm 4.3 is bounded as

$$R(t) \le \Delta \sum_{j=1}^{t} \mathbb{E}\big[\mathbb{I}[\mathcal{M}(j) \neq \mathcal{M}^*]\big] + \Delta \sum_{j=1}^{t} \mathbb{E}\Bigg[\mathbb{I}[\mathcal{M}(j) = \mathcal{M}^*] \times$$
$$\mathbb{I}\Big[\bigcup_{i<k, i\in\mathcal{I}_{\mathcal{M}^*}(j), k\in\mathcal{I}_{\mathcal{M}^*}(j)} \left\{\hat{\theta}_i(j) + \frac{1}{P_d - P_f}\sqrt{\frac{2\ln(j-1)}{T_i(j-1)}} < \hat{\theta}_k(j) + \frac{1}{P_d - P_f}\sqrt{\frac{2\ln(j-1)}{T_k(j-1)}}\right\}\Big]\Bigg] \tag{4.15}$$

where $\mathcal{I}_{\mathcal{M}^*}(j)$ denotes sensed-idle channels in Slot j when channels in $\mathcal{M}^*$ are sensed. On the right hand side of (4.15), the first term is the regret bound when the secondary user does not select exactly $\mathcal{M}^*$ to sense (i.e., $\mathcal{M}(j) \neq \mathcal{M}^*$), and the second term is the regret bound when the secondary user senses channels in $\mathcal{M}^*$ but does not select the best sensed-idle channel to access.

[5]This is consistent with the result in the subsequent Theorem 4.3 that the regret of Algorithm 4.3 is $O(\ln t)$.

In the sequel of this proof, for Slot j, denote $\hat{\theta}_k^T(T_k(j-1))$ as the estimated idle probability of Channel k, as described in Algorithm 4.3, when Channel k has been sensed by $T_k(j-1)$ slots until Slot $j-1$.

Now we derive a bound for the first term on the right hand side of (4.15). Recall that $T_i(t)$ is the number of slots in which Channel i is sensed until Slot t. Then we have

$$\sum_{j=1}^{t} \mathbb{E}\big[\mathbb{I}\left[\mathcal{M}(j) \neq \mathcal{M}^*\right]\big] \leq \sum_{i=M+1}^{N} \mathbb{E}[T_i(t)]. \tag{4.16}$$

Further, for $M+1 \leq i \leq N$ and any positive integer l, we have

$$\begin{aligned}
T_i(t) &= 1 + \sum_{j=\lceil \frac{N}{M} \rceil+1}^{t} \mathbb{I}\left[i \in \mathcal{M}(j)\right] \\
&= 1 + \sum_{j=\lceil \frac{N}{M} \rceil+1}^{t} \mathbb{I}\left[i \in \mathcal{M}(j),\, T_i(j-1) \geq l\right] + \sum_{j=\lceil \frac{N}{M} \rceil+1}^{t} \mathbb{I}\left[i \in \mathcal{M}(j),\, T_i(j-1) < l\right] \\
&\leq l + \sum_{j=\lceil \frac{N}{M} \rceil+1}^{t} \mathbb{I}\left[i \in \mathcal{M}(j), T_i(j-1) \geq l\right] \\
&\leq l + \sum_{j=\lceil \frac{N}{M} \rceil+1}^{t} \mathbb{I}\Bigg[\min_{k \in \mathcal{M}^*} \left\{\hat{\theta}_k^T(T_k(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_k(j-1)}}\right\} \\
&\qquad\qquad \leq \hat{\theta}_i^T(T_i(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_i(j-1)}},\, T_i(j-1) \geq l\Bigg] \\
&\leq l + \sum_{k=1}^{M} \sum_{j=\lceil \frac{N}{M} \rceil}^{t-1} \mathbb{I}\left[\hat{\theta}_k^T(T_k(j)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{T_k(j)}} \leq \hat{\theta}_i^T(T_i(j)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{T_i(j)}},\, T_i(j) \geq l\right] \\
&\leq l + \sum_{k=1}^{M} \sum_{j=\lceil \frac{N}{M} \rceil}^{t-1} \mathbb{I}\left[\min_{0<t_1 \leq j} \left\{\hat{\theta}_k^T(t_1) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_1}}\right\} \leq \max_{l \leq t_2 \leq j} \left\{\hat{\theta}_i^T(t_2) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_2}}\right\}\right] \\
&\leq l + \sum_{k=1}^{M} \sum_{j=1}^{t} \sum_{t_1=1}^{j} \sum_{t_2=l}^{j} \mathbb{I}\left[\hat{\theta}_k^T(t_1) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_1}} \leq \hat{\theta}_i^T(t_2) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_2}}\right].
\end{aligned} \tag{4.17}$$

Similar to analysis in [97], we have the fact that if event $\hat{\theta}_k^T(t_1) + \frac{1}{P_d-P_f}\sqrt{\frac{2\ln j}{t_1}} \leq \hat{\theta}_i^T(t_2) + \frac{1}{P_d-P_f}\sqrt{\frac{2\ln j}{t_2}}$ happens, then at least one of the following three events will happen: $\hat{\theta}_k^T(t_1) \leq \theta_k - \frac{1}{P_d-P_f}\sqrt{\frac{2\ln j}{t_1}}$, $\hat{\theta}_i^T(t_2) \geq \theta_i + \frac{1}{P_d-P_f}\sqrt{\frac{2\ln j}{t_2}}$, and $\theta_k < \theta_i + \frac{2}{P_d-P_f}\sqrt{\frac{2\ln j}{t_2}}$. In other words, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{I}\left[\hat{\theta}_k^T(t_1) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_1}} \leq \hat{\theta}_i^T(t_2) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_2}}\right]\right] \\
&\leq \mathbb{E}\left[\mathbb{I}\left[\hat{\theta}_k^T(t_1) \leq \theta_k - \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_1}}\right]\right] + \mathbb{E}\left[\mathbb{I}\left[\hat{\theta}_i^T(t_2) \geq \theta_i + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_2}}\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{I}\left[\theta_k < \theta_i + \tfrac{2}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_2}}\right]\right].
\end{aligned} \tag{4.18}$$

Using Chernoff-Hoeffding bound, the first two terms on the right hand side of (4.18) are bounded as

$$\mathbb{E}\Big[\mathbb{I}\big[\hat{\theta}_k^T(t_1) \le \theta_k - \frac{1}{P_d - P_f}\sqrt{\frac{2\ln j}{t_1}}\big]\Big] \le j^{-4}, \mathbb{E}\Big[\mathbb{I}\big[\hat{\theta}_i^T(t_2) \ge \theta_i + \frac{1}{P_d - P_f}\sqrt{\frac{2\ln j}{t_2}}\big]\Big] \le j^{-4}. \tag{4.19}$$

We note that if $t_2 \ge \frac{8\ln t}{(\theta_M-\theta_i)^2(P_d-P_f)^2}$, then we always have $\theta_k \ge \theta_i + \frac{2}{P_d-P_f}\sqrt{\frac{2\ln j}{t_2}}$ for any $k \in \mathcal{M}^*$ and $j \le t$, which means $\mathbb{I}\left[\theta_k < \theta_i + \frac{2}{P_d-P_f}\sqrt{\frac{2\ln j}{t_2}}\right] = 0$. Therefore, by setting $l = \left\lceil \frac{8\ln t}{(\theta_M-\theta_i)^2(P_d-P_f)^2} \right\rceil$, from (4.16)-(4.19) we have

$$\begin{aligned}
&\sum_{j=1}^{t} \mathbb{E}\left[\mathbb{I}\left[\mathcal{M}(j) \ne \mathcal{M}^*\right]\right] \\
&\quad\le \sum_{i=M+1}^{N} \left\lceil \frac{8\ln t}{(\theta_M-\theta_i)^2(P_d-P_f)^2} \right\rceil + \sum_{i=M+1}^{N}\sum_{k=1}^{M}\sum_{j=1}^{\infty}\sum_{t_1=1}^{j}\sum_{t_2=\left\lceil \frac{8\ln t}{(\theta_M-\theta_i)^2(P_d-P_f)^2}\right\rceil}^{j} 2j^{-4} \\
&\quad\le \sum_{i=M+1}^{N} \frac{8\ln t}{(\theta_M-\theta_i)^2(P_d-P_f)^2} + (N-M)\left(\frac{M\pi^2}{3}+1\right).
\end{aligned} \tag{4.20}$$

To bound the second term on the right hand side of (4.15), we have

$$\begin{aligned}
&\sum_{j=1}^{t} \mathbb{I}[\mathcal{M}(j) = \mathcal{M}^*]\mathbb{I}\Big[\bigcup_{i<k, i\in\mathcal{I}_{\mathcal{M}^*}(j), k\in\mathcal{I}_{\mathcal{M}^*}(j)} \big[\hat{\theta}_i^T(T_i(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_i(j-1)}} \\
&\qquad < \hat{\theta}_k^T(T_k(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_k(j-1)}}\big]\Big] \\
&\le 1 + \sum_{j=\lceil \frac{N}{M}\rceil+1}^{t} \mathbb{I}[\mathcal{M}(j) = \mathcal{M}^*]\mathbb{I}\Big[\bigcup_{i<k, i,k\in\mathcal{I}_{\mathcal{M}^*}(j)} \big[\hat{\theta}_i^T(T_i(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_i(j-1)}} \\
&\qquad < \hat{\theta}_k^T(T_k(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_k(j-1)}}\big]\Big] \\
&\le 1 + \sum_{i<k,\ i,k\in\mathcal{M}^*}\ \sum_{j=\lceil \frac{N}{M}\rceil+1}^{t} \mathbb{I}[\mathcal{M}(j) = \mathcal{M}^*]\mathbb{I}\Big[\hat{\theta}_i^T(T_i(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_i(j-1)}} \\
&\qquad < \hat{\theta}_k^T(T_k(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_k(j-1)}}\Big] \\
&\le \sum_{i<k,\ i,k\in\mathcal{M}^*} \Big\{ l_{i,k} + \sum_{j=\lceil \frac{N}{M}\rceil+1}^{t} \Big(\mathbb{I}\left[\mathcal{M}(j) = \mathcal{M}^*, T_k(j-1) \ge l_{i,k}\right] \\
&\qquad \cdot \mathbb{I}\left[\hat{\theta}_i^T(T_i(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_i(j-1)}} < \hat{\theta}_k^T(T_k(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_k(j-1)}}\right]\Big)\Big\} \\
&\le \sum_{i<k,\ i,k\in\mathcal{M}^*} \Big\{ l_{i,k} + \sum_{j=1}^{t}\sum_{t_1=1}^{j}\sum_{t_2=l_{i,k}}^{j} \mathbb{I}\left[\hat{\theta}_i^T(t_1) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_1}} < \hat{\theta}_k^T(t_2) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln j}{t_2}}\right]\Big\}
\end{aligned} \tag{4.21}$$

where $l_{i,k}$ can be an arbitrary positive integer.

Similar to the treatments in (4.18)-(4.20), the second term on the right hand side of

(4.15) is bounded as

$$\begin{aligned}\Delta \sum_{j=1}^{t} \mathbb{E}\Bigg[\mathbb{I}[\mathcal{M}(j) = \mathcal{M}^*]\mathbb{I}\Big[\bigcup_{i<k,i,k\in\mathcal{I}_{\mathcal{M}^*}(j)} &\{\hat{\theta}_i^T(T_i(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_i(j-1)}} \\ &< \hat{\theta}_k^T(T_k(j-1)) + \tfrac{1}{P_d-P_f}\sqrt{\tfrac{2\ln(j-1)}{T_k(j-1)}}\}\Big]\Bigg] \\ \leq \Delta \ln t \sum_{i<k\in\mathcal{M}^*} \tfrac{8}{(\theta_i-\theta_k)^2(P_d-P_f)^2} + \Delta\tbinom{M}{2}\big(\tfrac{\pi^2}{3}+1\big).\end{aligned} \tag{4.22}$$

Then, from (4.15), (4.20) and (4.22), the regret until Slot t, $R(t)$, is bounded as

$$\begin{aligned} R(t) \quad \leq \quad & \Delta \ln t \sum_{i=M+1}^{N} \tfrac{8}{(\theta_M-\theta_i)^2(P_d-P_f)^2} + \Delta \ln t \sum_{i<k\in\mathcal{M}^*} \tfrac{8}{(\theta_i-\theta_k)^2(P_d-P_f)^2} \\ & + \Delta(N-M)\left(\tfrac{M\pi^2}{3}+1\right) + \Delta\tbinom{M}{2}\left(\tfrac{\pi^2}{3}+1\right). \end{aligned} \tag{4.23}$$

In other words, $R(t) \sim O(\ln t)$ for finite t and for $t \to \infty$. $\square$

4.3.1.2 Multiple Channel Access at a Slot ($K > 1$)

When the secondary user can simultaneously access K channels at a slot, we modify Algorithm 4.3 as follows: in Line 6, instead of accessing a single channel, the secondary user selects up to K channels in $\mathcal{I}(t)$ with the largest values of $\hat{\theta}_i(t) + \frac{1}{P_d-P_f}\sqrt{\frac{2\ln(t-1)}{T_i(t-1)}}$. Similar to proof of Theorem 4.3, it can be proved that the regret of the resulted rule is $O(\ln t)$ for finite t and for $t \to \infty$.

4.3.2 Heterogenous Sensing

Consider that Channel i ($i = 1, ..., N$) has distinct setting $\left\{P_d^i, P_f^i\right\}$. The genie-aided rule with known channel statistics $\mathbf{\Theta}$ is still used as a benchmark of performance.

When channel statistics $\mathbf{\Theta}$ is unknown, it is desired to find a rule of good performance on regret $R(t)$ under heterogenous sensing. Then a question is raised: can we find a similar rule to those in Section 4.3.1, with $R(t) \sim O(\ln t)$ for finite t and for $t \to \infty$? To answer this question, we first look into the insights in the rules in Section 4.3.1.

As aforementioned, in Case II (partial channel sensing), there are two levels of MABP: the first level is to select which channels to sense, i.e., at Slot j, select channel set $\mathcal{M}$ to maximize

$$\mathbb{E}\big[\max_{\mathcal{K}(j)\subset\mathcal{I}_{\mathcal{M}}(j), |\mathcal{K}(j)|\leq K} \sum_{i\in\mathcal{K}(j)} \mathbb{E}[S_i(j)|X_i(j)=1]\big]$$

while the second level is to select which channels to access, i.e., to select sensed-idle channels with the largest $\mathbb{E}[S_i(j)|X_i(j) = 1]$. With homogeneous sensing, the criterion in the first level is simplified to finding the M channels with largest M θ_i's, while the criterion in the second level is simplified to, among sensed-idle channels, finding up to K channels with the largest θ_i's. Therefore, in Algorithm 4.3, in both levels we use sample mean of sensing observations of each channel to estimate θ_i. On the other hand, with heterogeneous sensing, the criteria in the two levels cannot be simplified to finding channels with the largest θ_i's. Therefore, it is not feasible to use sample mean of sensing observations as Algorithm 4.3 does. Rather, we need samples to reflect reward of each arm in each level, as shown in the following.

4.3.2.1 Single Channel Access at a Slot ($K = 1$)

Since the secondary user can sense M channels at a slot, the secondary user can sense one from $\binom{N}{M}$ possible sets of M channels, denoted $\mathcal{M}_1, \mathcal{M}_2, ..., \mathcal{M}_{\binom{N}{M}}$. In set $\mathcal{M}_i$ ($i = 1, 2, ..., \binom{N}{M}$), let $m_{i,j}$ ($j = 1, 2, ..., M$) denote the jth channel in $\mathcal{M}_i$. If the secondary user senses set $\mathcal{M}_i$ at Slot t, let $\mathcal{I}_{\mathcal{M}_i}(t)$ represent the sensing result, which is the set of sensed-idle channels. Until Slot t, let $T_i(t)$ denote the number of slots in which $\mathcal{M}_i$ is sensed, and $Y_i(t)$ denote the cumulative reward of the slots in which $\mathcal{M}_i$ is sensed. Until Slot t, let $T_{i,j}(t)$ ($j = 1, 2, ..., M$) denote the number of slots in which $\mathcal{M}_i$ is sensed and subsequently Channel $m_{i,j}$ is accessed, and $Y_{i,j}(t)$ denote the cumulative reward of Channel $m_{i,j}$ in slots in which $\mathcal{M}_i$ is sensed and subsequently Channel $m_{i,j}$ is accessed. Note that when we say "reward", it means the secondary user transmits over a channel, and receives acknowledgement (ACK) for the transmission. Here we assume the ACK transmission is error free. If no ACK is received (which means secondary transmission fails due to collision with primary transmission which is not detected), the reward of the corresponding secondary transmission is 0. The proposed OCSA rule is given in Algorithm 4.4. The secondary user keeps records of $T_i(t)$, $Y_i(t)$, $T_{i,j}(t)$, and $Y_{i,j}(t)$. In the sequel, for simplicity of presentation, the index (t) may be omitted for $T_i(t)$, $Y_i(t)$, $T_{i,j}(t)$, and $Y_{i,j}(t)$.

Remark 4.4. For set $\mathcal{M}_i$, $\frac{Y_i}{T_i}$ is the sample-mean of the reward if the channels in $\mathcal{M}_i$ are sensed. It is desired that a channel set with the largest reward is sensed. Similar to Algorithm 4.3, to guarantee each channel set is sensed sufficiently, we add an extra term, $\sqrt{\frac{2\ln(t-1)}{T_i}}$, to $\frac{Y_i}{T_i}$, to form the index in Line 4 of Algorithm 4.4. In Line 8 of Algorithm

Algorithm 4.4 Single Channel Access with Heterogeneous Sensing in Case II (Partial Channel Sensing)

---

for $i = 1 : \binom{N}{M}$ **do**
  Keep sensing $\mathcal{M}_i$ in continuous slots, and at each slot access one idle channel that was not accessed before when $\mathcal{M}_i$ is sensed. This procedure is repeated until each channel in $\mathcal{M}_i$ has been accessed at least once. For each slot, update T_i, Y_i, $T_{i,j}$, and $Y_{i,j}$, $j = 1, 2..., M$.
for each subsequent Slot t **do**
  Calculate indices $\frac{Y_i}{T_i} + \sqrt{\frac{2\ln(t-1)}{T_i}}$ ($i \in \{1, 2, ..., \binom{N}{M}\}$), and choose $i^\dagger = \arg\max_{i=1,...,\binom{N}{M}} \left\{\frac{Y_i}{T_i} + \sqrt{\frac{2\ln(t-1)}{T_i}}\right\}$.
  Sense channels in $\mathcal{M}_{i^\dagger}$
  if $\mathcal{I}_{\mathcal{M}_{i^\dagger}}(t)$, the set of sensed-idle channels at Slot t, is nonempty **then**
    Calculate indices $\frac{Y_{i^\dagger,j}}{T_{i^\dagger,j}} + \sqrt{\frac{2\ln(t-1)}{T_{i^\dagger,j}}}$, $m_{i^\dagger,j} \in \mathcal{I}_{\mathcal{M}_{i^\dagger}}(t)$.
    Select $j^\dagger = \arg\max_{m_{i^\dagger,j} \in \mathcal{I}_{\mathcal{M}_{i^\dagger}}(t)} \left\{\frac{Y_{i^\dagger,j}}{T_{i^\dagger,j}} + \sqrt{\frac{2\ln(t-1)}{T_{i^\dagger,j}}}\right\}$, access Channel $m_{i^\dagger,j^\dagger}$, and check whether the transmission is successful.
    Update $T_{i^\dagger}$, $Y_{i^\dagger}$, $T_{i^\dagger,j^\dagger}$, $Y_{i^\dagger,j^\dagger}$.
  else
    Update $T_{i^\dagger}$.

4.4, the extra term $\sqrt{\frac{2\ln(t-1)}{T_{i^\dagger,j}}}$ is to guarantee each channel in $\mathcal{M}_{i^\dagger}$, say Channel $m_{i^\dagger,j}$, is accessed in a sufficient number of slots such that, after Channel $m_{i^\dagger,j}$ has been sufficiently sensed, the sample mean $\frac{Y_{i^\dagger,j}}{T_{i^\dagger,j}}$, which dominates in the index when $T_{i^\dagger,j}$ is large, can accurately represent the real average reward of Channel $m_{i^\dagger,j}$ at a slot.

Theorem 4.4. The regret $R(t)$ of Algorithm 4.4 is $O(\ln t)$ with $t \to \infty$ and with finite t.

Proof. Denote $\mathcal{M}_{i^*}$ as the optimal set of channels to sense (i.e., the set of channels to sense in the genie-aided rule). Denote $\mathcal{M}(t)$ as the channel set decided by Algorithm 4.4 to be sensed at Slot t. Similar to proof of Theorem 4.3, the regret $R(t)$ until Slot t is bounded as

$$
\begin{aligned}
&R(t) \leq \\
&\Delta \sum_{j=1}^{t} \mathbb{E}\big[\mathbb{I}[\mathcal{M}(j) \neq \mathcal{M}_{i^*}]\big] + \Delta \sum_{j=1}^{t} \mathbb{E}\Bigg[\mathbb{I}[\mathcal{M}(j) = \mathcal{M}_{i^*}] \times \mathbb{I}\Big[\bigcup_{\substack{m_{i^*,k}, m_{i^*,r} \in \mathcal{I}_{\mathcal{M}_{i^*}}(j) \\ \mathbb{E}[S_{m_{i^*,k}}|X_{m_{i^*,k}}=1] > \mathbb{E}[S_{m_{i^*,r}}|X_{m_{i^*,r}}=1]}} \\
&\qquad \Big\{\frac{Y_{i^*,k}(j-1)}{T_{i^*,k}(j-1)} + \sqrt{\frac{2\ln(j-1)}{T_{i^*,k}(j-1)}} < \frac{Y_{i^*,r}(j-1)}{T_{i^*,r}(j-1)} + \sqrt{\frac{2\ln(j-1)}{T_{i^*,r}(j-1)}}\Big\}\Big]\Bigg]. \qquad (4.24)
\end{aligned}
$$

Next we derive bounds for the two terms on the right hand side of (4.24), respectively.

Since $T_i(t)$ is the number of slots that channel set $\mathcal{M}_i$ is sensed until Slot t, the first term on the right hand side of (4.24) is $\Delta \sum\limits_{j=1}^{t} \mathbb{E}\left[\mathbb{I}\left[\mathcal{M}(j) \neq \mathcal{M}_{i^*}\right]\right] = \Delta \sum\limits_{i \neq i^*, i \in \{1,2,...,\binom{N}{M}\}} \mathbb{E}[T_i(t)]$.

For each $i \in \{1, 2, ..., \binom{N}{M}\}$, it can be proved that the reward sequence $Y_i(t)|_{T_i(t)=1}$, $Y_i(t)|_{T_i(t)=2}$, ..., $Y_i(t)|_{T_i(t)=n}$ satisfies a so-called *drift condition*[6], with the proof given in Appendix 4.B.

Similar to the treatments in (4.17)-(4.20), we have $\mathbb{E}[T_i(t)] \leq \frac{8 \ln t}{\xi_i} + \frac{\pi^2}{3} + 1$ where

$$\xi_i \triangleq \Big(\mathbb{E}\Big[\max_{l \in \mathcal{I}_{\mathcal{M}_{i^*}}} \mathbb{E}\left[S_l | X_l = 1\right]\Big] - \mathbb{E}\Big[\max_{l \in \mathcal{I}_{\mathcal{M}_i}} \mathbb{E}\left[S_l | X_l = 1\right]\Big]\Big)^2$$

and $\mathcal{I}_{\mathcal{M}_i}$ is the set of sensed-idle channels if $\mathcal{M}_i$ is sensed. Therefore, the first term on the right hand side of (4.24) is bounded as

$$\Delta \sum_{j=1}^{t} \mathbb{E}\big[\mathbb{I}[\mathcal{M}(j) \neq \mathcal{M}_{i^*}]\big] \leq \Delta \ln t \sum_{\substack{i \in \{1,2,...,\binom{N}{M}\} \\ i \neq i^*}} \frac{8}{\xi_i} + \Delta(\binom{N}{M} - 1)(\frac{\pi^2}{3} + 1). \quad (4.25)$$

Similar to the treatments in (4.21)-(4.22), we have a bound for the second term on the right hand side of (4.24) as $\Delta \ln t \sum\limits_{k<r \leq M} \frac{8}{\left(\frac{(1-P_f^{m_{i^*,k}})\theta_{m_{i^*,k}}}{f(\theta_{m_{i^*,k}})} - \frac{(1-P_f^{m_{i^*,r}})\theta_{m_{i^*,r}}}{f(\theta_{m_{i^*,r}})}\right)^2} + \Delta\binom{M}{2}\big(\frac{\pi^2}{3} + 1\big)$.

It can be seen that, the two terms on the right hand side of (4.24) are both bounded by $O(\ln t)$. Therefore, the regret until Slot t, $R(t)$, is $O(\ln t)$. □

4.3.2.2 Multiple Channel Access at a Slot ($K > 1$)

When the secondary user can simultaneously access up to K channels at a slot, we modify Algorithm 4.4 as follows: In Lines 8 and 9, the secondary user selects to access up to K sensed-idle channels with the largest values of $\frac{Y_{i^\dagger,j}}{T_{i^\dagger,j}} + \sqrt{\frac{2\ln(t-1)}{T_{i^\dagger,j}}}$, $m_{i^\dagger,j} \in \mathcal{I}_{\mathcal{M}_{i^\dagger}}(t)$, and updates $T_{i^\dagger,j}$ and $Y_{i^\dagger,j}$ accordingly if Channel $m_{i^\dagger,j}$ is accessed. Similarly, it can be proved that the regret of the resulted rule is $O(\ln t)$ with finite t and with $t \to \infty$.

[6]Its definition is given in Section 2.4 of [98].

TABLE 4.1
PARAMETERS USED IN THE SIMULATION.

$N=5$	P_d^i	heter.: (0.9286,0.7729, 0.8878, 0.9627, 0.9364)
	P_f^i	heter.: (0.2292, 0.3031, 0.0829, 0.2335, 0.1573)
	θ_i (i.i.d.)	(0.5296,0.4001,0.9817,0.1931,0.2495)
$N=6$	P_d^i	heter.: (0.8961,0.8298, 0.7675, 0.7529, 0.8331, 0.9167)
	P_f^i	heter.: (0.1653, 0.2700, 0.1448, 0.0777, 0.1093,0.0341)
	θ_i (i.i.d.)	(0.3605,0.9291,0.7694,0.6199,0.4109,0.3559)
$N=7$	P_d^i	heter.: (0.8700,0.7788, 0.8595, 0.8134, 0.8958, 0.9128, 0.8932)
	P_f^i	heter.: (0.1567, 0.3204, 0.2472, 0.3354, 0.3458, 0.3093, 0.0899)
	θ_i (i.i.d.)	(0.8811,0.5390,0.3468,0.9522,0.7823,0.0471,0.7968)
$N=8$	P_d^i	heter.: (0.8556,0.7283,0.9319,0.7260,0.8103,0.8707,0.9165,0.8359)
	P_f^i	heter.: (0.2377,0.2342,0.2552,0.1397,0.1350,0.1699,0.3101,0.1497)
	θ_i (i.i.d.)	(0.6923,0.5430,0.3544,0.8753,0.5212,0.6759,0.8783,0.9762)

4.4 Performance Evaluation

We use Monte-Carlo simulation to validate our analysis. Consider a CRN with $N = 5, 6, 7, 8$ channels. Other parameters are listed in Table 4.1, where "heter." means heterogeneous sensing. And for homogeneous case with all N values, $P_d^i = 0.8$ and $P_f^i = 0.3$.

4.4.1 Full channel sensing with i.i.d. model

Case I (full channel sensing) with i.i.d. model is evaluated first. Figs. 4.1 and 4.2 show the average regret of Algorithm 4.1 with homogeneous sensing and heterogeneous sensing, respectively. It can be seen that when t is large, the regret $R(t)$ tends to be finitely bounded. Figs. 4.3 and 4.4 show the average regret of Algorithm 4.2 with homogeneous sensing and heterogeneous sensing, respectively. The regret $R(t)$ also tends to be finitely bounded. Fig. 4.5 shows the impact of K (the maximal number of channels that can be accessed) in Algorithm 4.2 with homogeneous sensing. It can be seen that $R(t)$ increases when K changes from 1 to 3, and $R(t)$ decreases when K further changes to 5 and 7. This can be explained as follows. When $K = 1$, the wrong access (i.e., the proposed rule does not access the same channel as the genie-aided rule does) is only on one single channel. When K changes to 3, the wrong access is on up to 3 channels, and thus, the regret is likely to be larger than that with $K = 1$. When K further increases, the up-to-K channels selected by the proposed rule and the up-to-K channels selected by the genie-aided rule are likely to be

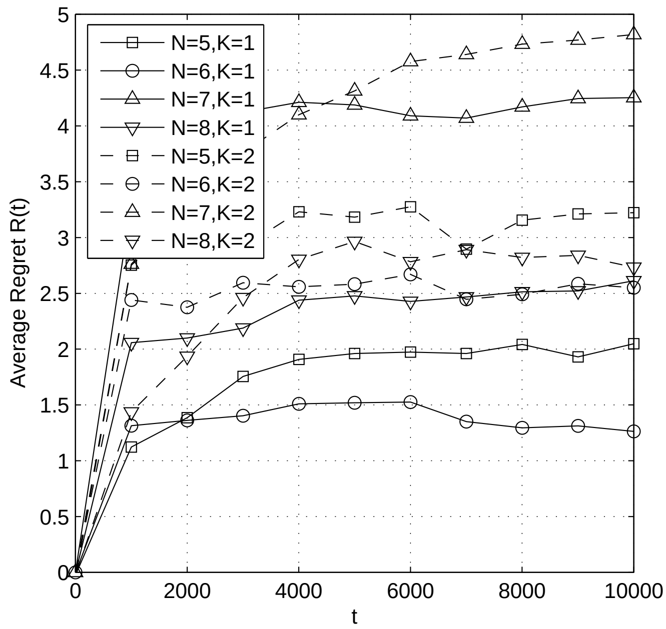

Fig. 4.1. Average regret $R(t)$ of Algorithm 4.1 with homogeneous sensing in Case I (full channel sensing), i.i.d. model.

with minor difference, and thus, the regret is reduced. As an extreme case, when $K = 8$ in this example (not shown in the figure), there is no difference between the channels selected by the proposed rule and the channels selected by the genie-aided rule, which means the regret is 0.

4.4.2 Partial channel sensing with i.i.d. model

Case II (partial channel sensing) with i.i.d. model is then evaluated. Since the proposed algorithms for partial channel sensing have regrets bounded by $O(\ln t)$, we evaluate normalized regret, given as $R(t)/\ln t$. Figs. 4.6 and 4.7 show average $R(t)/\ln t$ of the proposed algorithms in homogeneous sensing and heterogeneous sensing, respectively. It can be seen that the normalized regrets in the two figures are finitely bounded, which is consistent with our claim that the proposed algorithms have regrets bounded by $O(\ln t)$.

The impact of K in homogeneous sensing and partial channel sensing with the i.i.d. model is shown in Fig. 4.8, while the impact of K in heterogenous sensing and partial channel sensing with the i.i.d. model can be observed from Fig. 4.7. It can be seen that, for homogeneous sensing and partial channel sensing with the i.i.d. model, the normalized

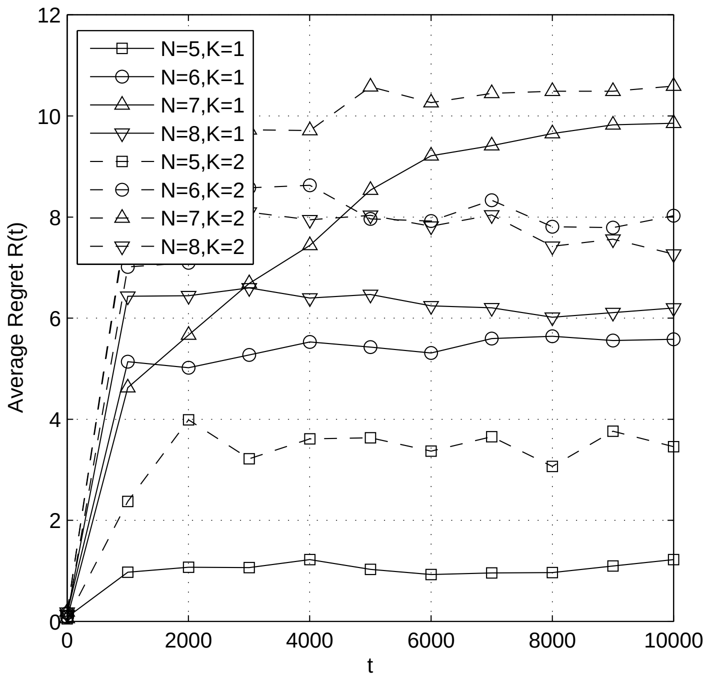

Fig. 4.2. Average regret $R(t)$ of Algorithm 4.1 with heterogeneous sensing in Case I (full channel sensing), i.i.d. model.

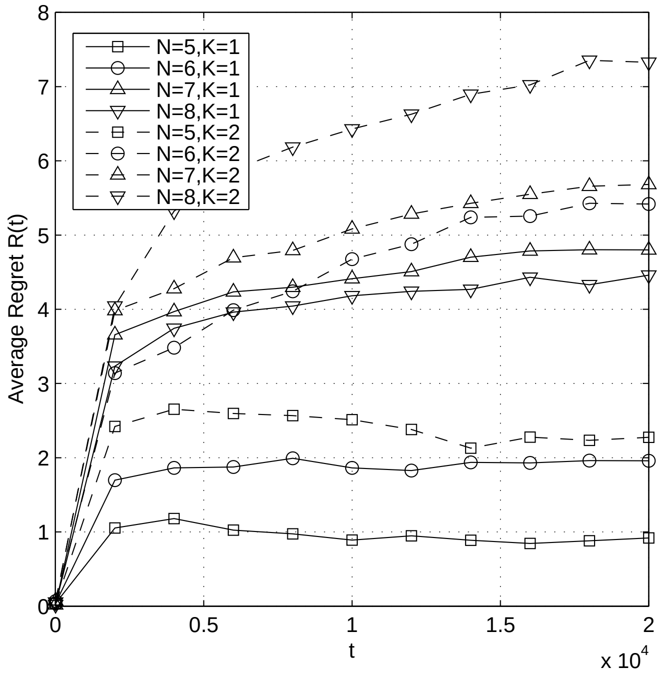

Fig. 4.3. Average regret $R(t)$ of Algorithm 4.2 with homogeneous sensing in Case I (full channel sensing), i.i.d. model.

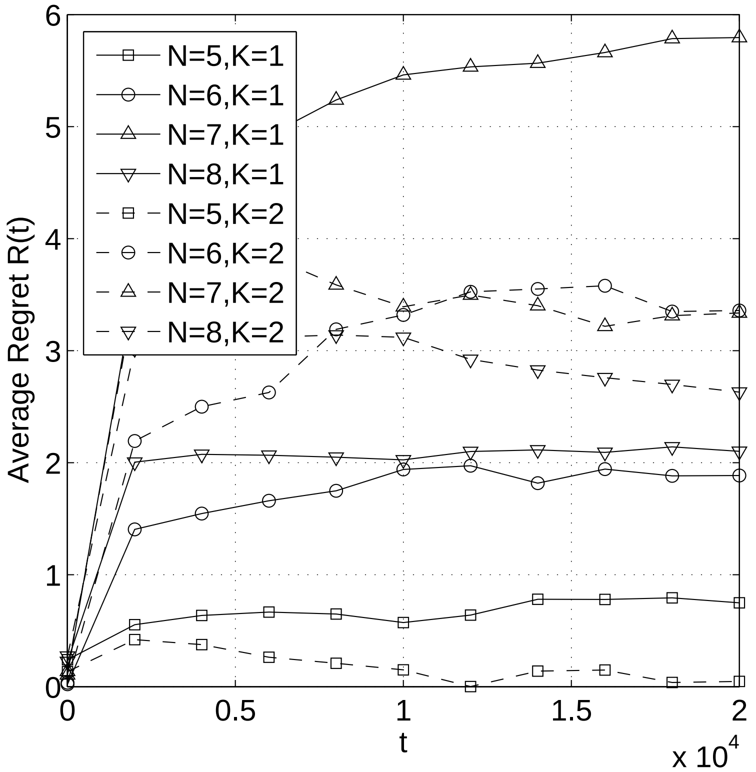

Fig. 4.4. Average regret $R(t)$ of Algorithm 4.2 with heterogeneous sensing in Case I (full channel sensing), i.i.d. model.

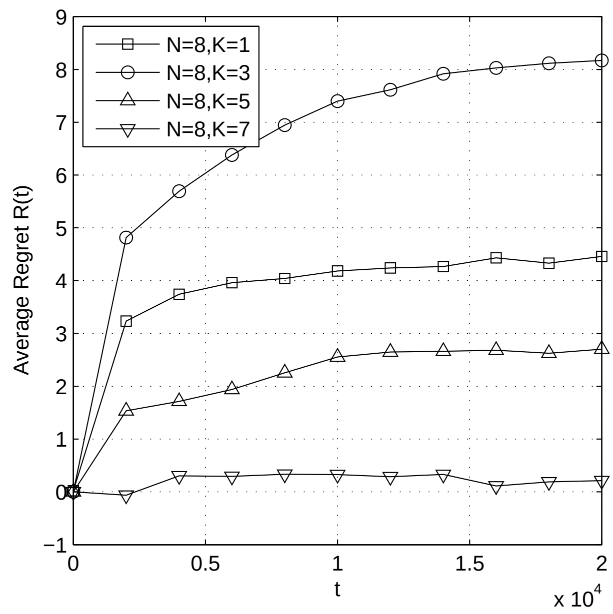

Fig. 4.5. The impact of K on the average regret $R(t)$ of Algorithm 4.2 with homogeneous sensing in Case I (full channel sensing), i.i.d. model.

regret increases when K increases from 1 to M. This is because, in partial channel sensing, the M sensed channels in our algorithms and in the genie-aided rule may not be the same. If we select more channels (larger K) to access, then the channels selected by our algorithms may have bigger difference from the channels selected by the genie-aided rule, and thus, the normalized regret increases. However, for heterogenous sensing and partial channel sensing with the i.i.d. model, from Fig. 4.7 it can be seen that the normalized regret decreases when K increases to M. This is because we use different algorithms in homogeneous sensing and heterogeneous sensing. In homogeneous sensing with the i.i.d. model, we use sensing results to update our indices (the selection metrics), while in heterogeneous sensing with the i.i.d. model, we use channel access results to update our indices. So in heterogeneous sensing with the i.i.d. model, when K is larger, the secondary user can access more channels, and can have more accurate estimation, and thus, has a larger chance to select the same set of channels to sense as the genie-aided rule, leading to smaller normalized regret.

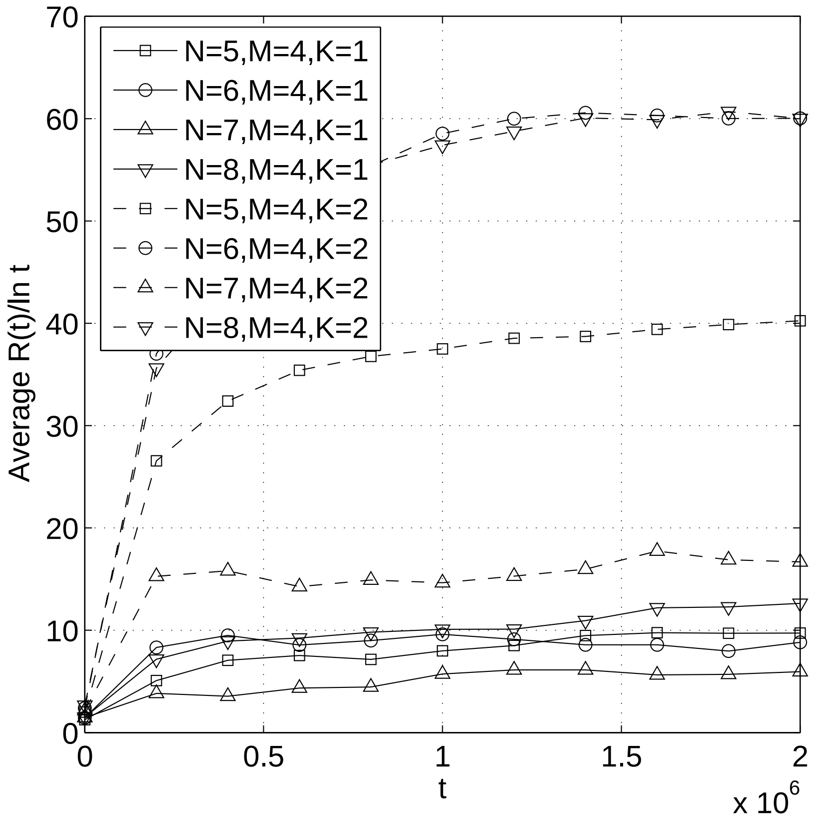

Fig. 4.6. Average $R(t)/\ln t$ of the proposed algorithm with homogeneous sensing in Case II (partial channel sensing), i.i.d. model.

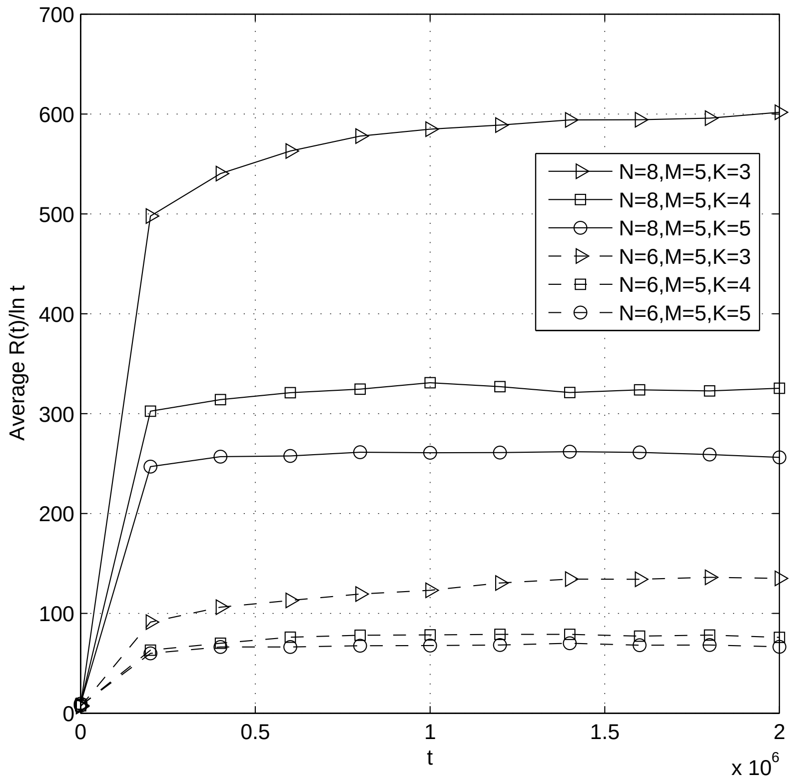

Fig. 4.7. Average $R(t)/\ln t$ of the proposed algorithm with heterogeneous sensing in Case II (partial channel sensing), i.i.d. model.

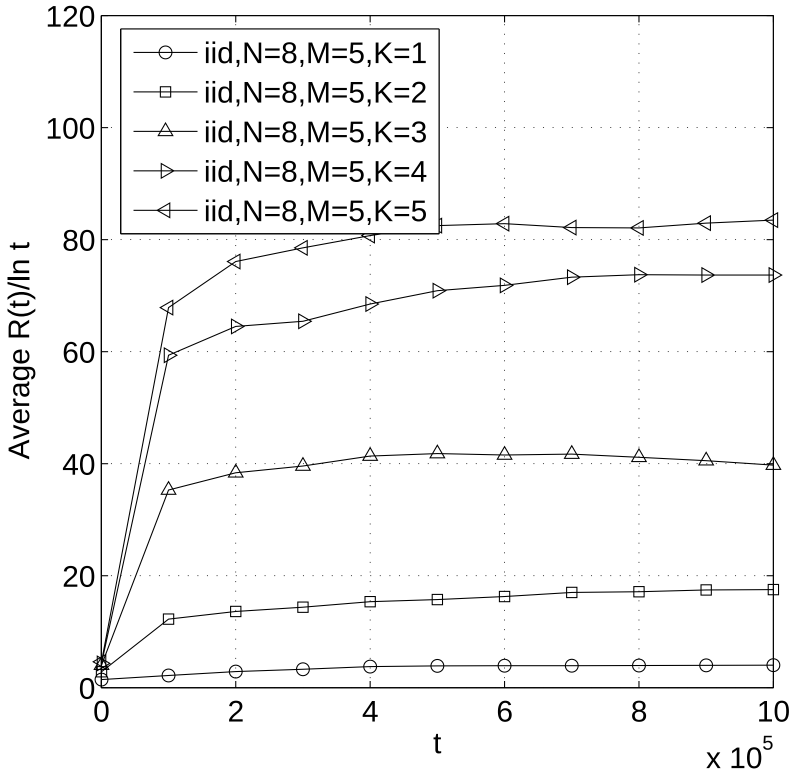

Fig. 4.8. Impact of K on the average $R(t)/\ln t$ of the proposed algorithms in homogeneous sensing and partial channel sensing with the i.i.d. model.

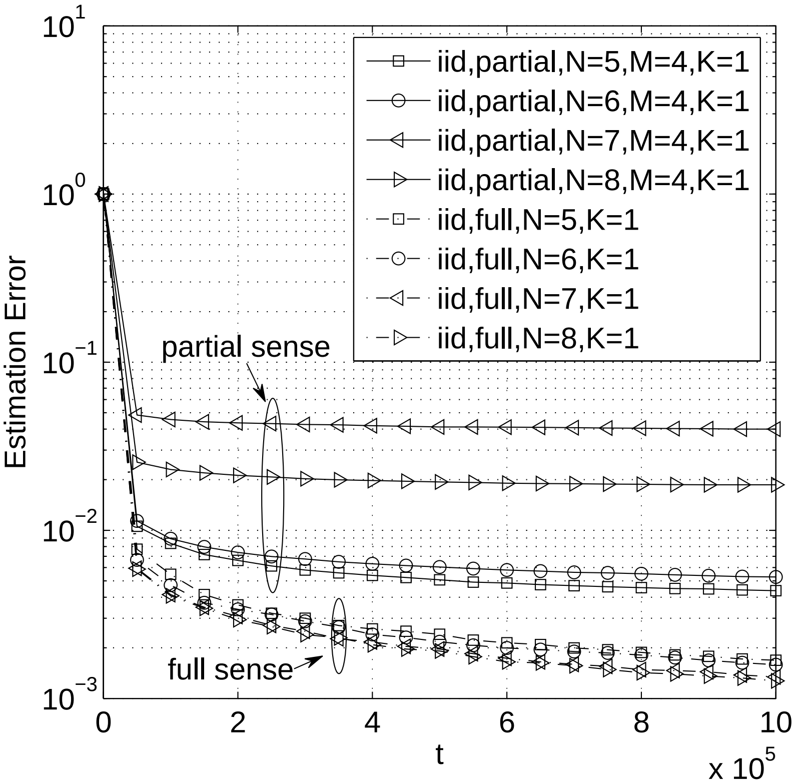

Fig. 4.9. Estimation errors of $\mathbf{\Theta} = (\theta_1, \theta_2, ..., \theta_N)$, i.i.d. model.

4.4.3 Estimation for θ_i's

Now we evaluate the accuracy of the estimators for channel idle probabilities θ_i's in the i.i.d. model. Fig. 4.9 shows the estimation errors for the i.i.d. model in full channel sensing of Algorithm 2 (denoted as "full"), i.i.d. model in homogeneous sensing and partial channel sensing (denoted as "iid")[7]. Here the estimation error is defined as ratio of L_2-norm of $(\mathbf{\Theta} - \hat{\mathbf{\Theta}})$ to L_2-norm of $\mathbf{\Theta}$ (i.e., $\frac{||\mathbf{\Theta}-\hat{\mathbf{\Theta}}||_2}{||\mathbf{\Theta}||_2}$) where $\mathbf{\Theta}$ and $\hat{\mathbf{\Theta}}$ are the vector of real channel idle probabilities and the vector of estimated channel idle probabilities, respectively. It can be seen that the estimation error tends to converge to 0 when t keeps increasing. This can be explained as follows. In full channel sensing, all channels are sensed at each time slot. So the estimation error should converge to 0 when $t \to \infty$, according to the Strong Law of Large Number. For partial channel sensing in Fig. 4.9, as explained in the Remark for

[7]For i.i.d. model in heterogeneous sensing and partial channel sensing, our proposed algorithm is not based on estimation of channel idle probabilities θ_i's. So we do not have results for accuracy of estimators for θ_i's.

Algorithm 4.3, the extra term in the indices determines that when t increases, the secondary user mainly senses the best channel (the channel with the largest θ_i), while other inferior channels will still be sensed in the scale of $O(\ln t)$. Therefore, the estimation error should also converge to 0 when $t \to \infty$, although the convergence speed is much smaller than that in the full channel sensing case.

4.4.4 Comparison with other schemes

Next we compare our proposed rules with other rules in the literature. Since reference [12] is one most related work, we compare our proposed rules with Rule 2 in [12], for partial channel sensing and homogeneous/heterogeneous sensing. The average normalized regrets in those rules are shown in Fig. 4.10. It can be seen that our proposed rules in homogeneous sensing and heterogeneous sensing have better performance than the rule in [12].

4.5 Conclusion

In this chapter, the problem of OCSA by a secondary user in a CRN is investigated. The i.i.d. model is investigated. In the case with full channel sensing, with side information through sensing all the channels, the regret due to unknown primary users' statistical information is proved to be asymptotically finite. On the other hand, for the case with partial channel sensing, asymptotically finite regret cannot be achieved since it is proved that the regret is at least $O(\ln t)$. Therefore, in our research we derive OCSA rules with regret $O(\ln t)$, for homogeneous sensing and heterogeneous sensing, respectively. This research should provide insights to the design of OCSA in CRNs with unknown statistical information of primary channels. Further research may include the case with competition among multiple secondary users, the generalization of our solution in i.i.d. model with partial channel sensing and heterogeneous sensing to a more general bi-level MABP, and the case with time-varying θ_i's.

Appendix 4.A: Proof of $\varepsilon > 0$

We have $\varepsilon \geq 0$. Next we prove $\varepsilon > 0$ by using proof by contradiction.

Suppose $\varepsilon = 0$. It means that, for any small enough $\sigma > 0$, there always exists $\mathbf{\Theta}^\dagger \in \mathcal{C}_e$

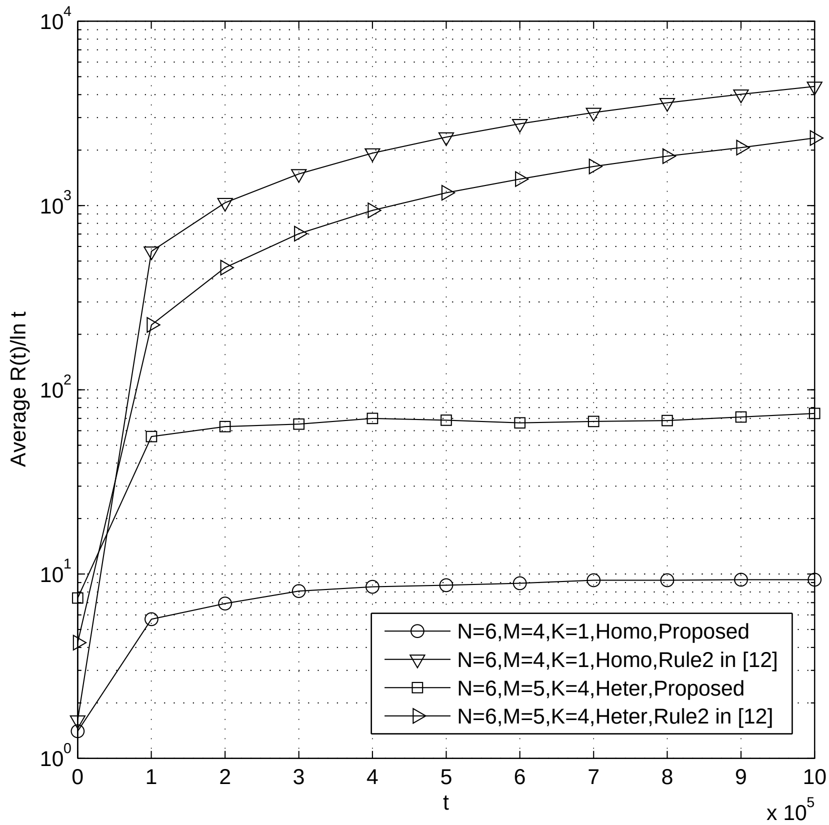

Fig. 4.10. Average $R(t)/\ln t$ of the proposed algorithms and Rule 2 in [12], with i.i.d. model, partial channel sensing, and homogeneous (denoted as “Homo’)/ heterogeneous (denoted as “Heter") sensing.

such that $||\mathbf{P}_{\mathbf{\Theta}^\dagger} - \mathbf{P}_{\mathbf{\Theta}}||_2 < \sigma$. In addition, it can be seen that $||\mathbf{P}_{\mathbf{\Theta}'} - \mathbf{P}_{\mathbf{\Theta}}||_2 < \sigma$ is a continuous function over $\mathbf{\Theta}'$, and equation $||\mathbf{P}_{\mathbf{\Theta}'} - \mathbf{P}_{\mathbf{\Theta}}||_2 = 0$ has a unique root $\mathbf{\Theta}' = \mathbf{\Theta}$.

Based on the above two facts, it can be concluded that $\mathbf{\Theta}^\dagger$ should approach sufficiently close to $\mathbf{\Theta}$.

On the other hand, since $\mathbf{\Theta}^\dagger \in \mathcal{C}_e$, there exists $\delta_0 > 0$ and $i \in \{1, ..., N\}$ such that $|\theta_i^\dagger - \theta_i| > \delta_0$,[8] which contradicts the above conclusion that $\mathbf{\Theta}^\dagger$ should approach sufficiently close to $\mathbf{\Theta}$. Thus it concludes $\varepsilon > 0$.

Appendix 4.B: Proof of Drift Condition

The drift condition is satisfied if 1) $\lim\limits_{n\to\infty} \mathbb{E}[\frac{Y_i(t)}{n}|T_i(t) = n] = \mathbb{E}\Big[\max\limits_{l\in\mathcal{I}_{\mathcal{M}_i}} \mathbb{E}\left[S_l|X_l = 1\right]\Big]$; 2) there exists n_0 such that if $n \geq n_0$, we have

$$\mathbb{P}\left(\frac{Y_i(t)|_{T_i(t)=n}}{n} \geq \mathbb{E}\Big[\frac{Y_i(t)}{n}\Big] + \sqrt{\frac{2\ln t}{n}}\right) \leq t^{-4} \tag{4.26}$$

$$\mathbb{P}\left(\frac{Y_i(t)|_{T_i(t)=n}}{n} \leq \mathbb{E}\Big[\frac{Y_i(t)}{n}\Big] - \sqrt{\frac{2\ln t}{n}}\right) \leq t^{-4}. \tag{4.27}$$

Part 2) is a conclusion directly from Theorem 4 in [98]. Next we give proof of 1).

Each channel set $\mathcal{M}_i$ has M channels, and thus, if set $\mathcal{M}_i$ is sensed, we have 2^M possible results of the set of sensed-idle channels, denoted $\mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_{2^M} \subseteq \{1, 2, ..., M\}$.[9] Denote $T_{i,k}$ as the number of time slots in which channel set $\mathcal{M}_i$ is sensed, and sensing result is $\mathcal{R}_k$. Denote $T_{i,k,j}$ as the number of time slots in which channel set $\mathcal{M}_i$ is sensed,

[8] Otherwise, no wrong access is made based on $\mathbf{\Theta}^\dagger$, and thus, $\mathbf{\Theta}^\dagger \notin \mathcal{C}_e$.

[9] As an example, if $\mathcal{R}_1 = \{1\}$, it means the first channel in $\mathcal{M}_i$ is sensed idle, and other channels in $\mathcal{M}_i$ are sensed busy.

sensing result is $\mathcal{R}_k$, and the jth channel in $\mathcal{M}_i$ is accessed.

$$\begin{aligned}
&\left|\mathbb{E}\Big[\max_{l\in\mathcal{I}_{\mathcal{M}_i}}\mathbb{E}\left[S_l|X_l=1\right]\Big]-\mathbb{E}[\tfrac{Y_i}{n}|T_i=n]\right|\\
=\;&\Big|\sum_{k=1}^{2^M}\mathbb{P}\big(\mathcal{R}_k\big)\max_{l\in\mathcal{R}_k}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}-\sum_{j=1}^{M}\frac{\mathbb{E}\Big[\sum\limits_{k=1}^{2^M}T_{i,k,j}\Big]}{n}\frac{(1-P_f^{m_{i,j}})\theta_{m_{i,j}}}{f(\theta_{m_{i,j}})}\Big|\\
=\;&\left|\sum_{k=1}^{2^M}\frac{\sum\limits_{j=1}^{M}\mathbb{E}[T_{i,k,j}]}{n}\max_{l\in\mathcal{R}_k}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}-\sum_{k=1}^{2^M}\frac{\sum\limits_{j=1}^{M}\mathbb{E}[T_{i,k,j}]}{n}\frac{(1-P_f^{m_{i,j}})\theta_{m_{i,j}}}{f(\theta_{m_{i,j}})}\right|\\
=\;&\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\in\mathcal{R}_k}\mathbb{E}[T_{i,k,j}]\max_{l\in\mathcal{R}_k}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}-\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\in\mathcal{R}_k}\mathbb{E}\big[T_{i,k,j}\big]\frac{(1-P_f^{m_{i,j}})\theta_{m_{i,j}}}{f(\theta_{m_{i,j}})}\\
\overset{(a)}{=}\;&\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}\mathbb{E}\left[T_{i,k,j}\right]\max_{l\in\mathcal{R}_k}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}-\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}\mathbb{E}\big[T_{i,k,j}\big]\frac{(1-P_f^{m_{i,j}})\theta_{m_{i,j}}}{f(\theta_{m_{i,j}})}\\
\le\;&\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}\mathbb{E}\left[T_{i,k,j}\right]\max_{l\in\mathcal{R}_k}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}\\
\le\;&\max_{l\in\mathcal{M}_i}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}\mathbb{E}\left[T_{i,k,j}\right]
\end{aligned}\tag{4.28}$$

where the first equality comes from Wald's lemma (see, e.g., Lemma 3.1 in [99]), and in (a) we have $j^*_{i,k}=\arg\max_{j\in\mathcal{R}_k}\frac{\left(1-P_f^{m_{i,j}}\right)\theta_{m_{i,j}}}{f(\theta_{m_{i,j}})}$ (in other words, when the channel set $\mathcal{M}_i$ is sensed, and sensing result is $\mathcal{R}_k$, then the optimal channel to be accessed is the $j^*_{i,k}$th channel in $\mathcal{M}_i$).

Given $T_{i,k}$ (the number of time slots in which channel set $\mathcal{M}_i$ is sensed, and sensing result is $\mathcal{R}_k$), the number of slots that the jth ($j\neq j^*_{i,k}$) channel in $\mathcal{M}_i$ is accessed has an expectation $\mathbb{E}[T_{i,k,j}|T_{i,k}]\le\frac{8\ln T_{i,k}}{\delta_{i,j,k}}+1+\frac{\pi^2}{3}$, where $\delta_{i,j,k}=\Big(\max_{l\in\mathcal{R}_k}\frac{\left(1-P_f^{m_{i,l}}\right)\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}-\frac{\left(1-P_f^{m_{i,j}}\right)\theta_{m_{i,j}}}{f(\theta_{m_{i,j}})}\Big)^2$. The proof is similar to (4.18)-(4.22).

Therefore, from (4.28) we have

$$\begin{aligned}
&\left|\mathbb{E}\Big[\max_{l\in\mathcal{I}_{\mathcal{M}_i}}\mathbb{E}\left[S_l|X_l=1\right]\Big]-\mathbb{E}[\tfrac{Y_i}{n}|T_i=n]\right|\\
\le\;&\max_{l\in\mathcal{M}_i}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}\left[\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}\mathbb{E}\Big[\frac{8\ln T_{i,k}}{\delta_{i,j,k}}\Big]+\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}(1+\tfrac{\pi^2}{3})\right]\\
\le\;&\max_{l\in\mathcal{M}_i}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}\left[\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}\frac{8}{\delta_{i,j,k}}\ln\mathbb{E}\Big[T_{i,k}\Big]+\frac{1}{n}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}(1+\tfrac{\pi^2}{3})\right]\\
\overset{(b)}{\le}\;&\frac{\ln n}{n}\left[\max_{l\in\mathcal{M}_i}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}\frac{8}{\delta_{i,j,k}}\right]\\
&+\frac{1}{n}\left[\max_{l\in\mathcal{M}_i}\frac{(1-P_f^{m_{i,l}})\theta_{m_{i,l}}}{f(\theta_{m_{i,l}})}\sum_{k=1}^{2^M}\sum_{j\neq j^*_{i,k},j\in\mathcal{R}_k}(1+\tfrac{\pi^2}{3})\right]
\end{aligned}\tag{4.29}$$

where (b) comes from $T_{i,k} \leq n$.

On the right hand side of the last inequality in (4.29), the first term is $\frac{\ln n}{n}$ times a constant, and the second term is $\frac{1}{n}$ times a constant. Therefore, when $n \to \infty$, both terms converge to 0. So from (4.29) we have

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{Y_i(t)}{n}|T_i(t)=n\right] = \mathbb{E}\Big[\max_{l\in\mathcal{I}_{\mathcal{M}_i}} \mathbb{E}\left[S_l|X_l=1\right]\Big],\ \forall i \in \left\{1,2,...,\binom{N}{M}\right\}. \quad (4.30)$$

Chapter 5

Channel Exploration and Exploitation with Imperfect Spectrum Sensing in CRNs: Markov Model of Channel Busy/Idle States over Time

In this chapter, the online learning problem of OCSA in CRNs is studied. The channel sensing is imperfect. A secondary user cannot sense all potential channels simultaneously, and can access up to a limited number of channels at a time. Different from Chapter 4, the channel busy/idle states of each channel over time slots follow a discrete-time Markov chain. The parameters of the Markov chain are unknown by the secondary user. In this research, the OCSA problem is modeled as restless multi-arm bandit problem. And an OCSA rule with logarithmic regret is derived.

5.1 System Model

In Chapter 4, for each channel, it has the same probability to be idle at all slots, and its channel busy/idle states over time slots follow the i.i.d. model. In this chapter, we consider that the channel busy/idle states of each channel over time slots follow a discrete-time Markov

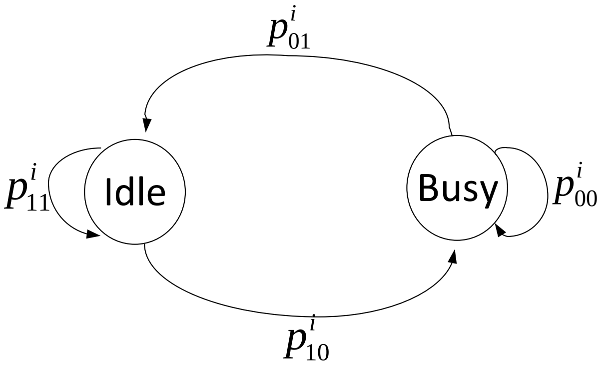

Fig. 5.1. The Markov model of busy/idle states of a channel (say Channel i) over time slots.

chain, i.e., with the Markov model as shown in Fig. 5.1. For Channel i at Slot j, $S_i(j) = 1$ means the channel is idle, and $S_i(j) = 0$ means the channel is busy. From one time slot to the next, the transition probability matrix is given as

$$\begin{pmatrix} p_{00}^i & 1 - p_{00}^i \\ 1 - p_{11}^i & p_{11}^i \end{pmatrix}$$

where p_{00}^i and p_{11}^i are the transition probability from a busy state to a busy state, and from an idle state to an idle state, respectively, and $p_{00}^i, p_{11}^i \in (0, 1)$. Therefore, the stationary idle probability of Channel i is given as $\theta_i = \frac{1-p_{00}^i}{2-p_{00}^i-p_{11}^i}$. All the N Markov chains of the N channels are independent. Here we consider homogeneous sensing with partial channel sensing (which is more general than the full channel sensing) and single channel access. Extension to multiple channel access is straightforward. The heterogeneous sensing case is more complicated, and will be investigated in our future work.

For performance comparison, it is desired to have the performance of a genie-aided rule as a benchmark. However, it is still unclear what the optimal performance of a genie-aided rule is when the secondary user does not access all sensed-idle channels with imperfect sensing. Therefore, similar to [100] and [101], a genie-aided rule is set up as: select the M channels with the highest stationary idle probabilities, and access the sensed-idle channel with the largest conditional reward. So for the genie-aided rule, the reward until Slot t is given as $t \cdot \mathbb{E}\big[\max_{i \in \mathcal{M}^*} \mathbb{E}[S_i | X_i = 1]\big]$ where $\mathcal{M}^*$ consists of the M channels with the largest θ_i, and the conditional reward of Channel i (conditioned on that Channel i is sensed idle) is given as $\mathbb{E}[S_i|X_i = 1] = \frac{(1-P_f^i)\theta_i}{f(\theta_i)}$.

For a practical OCSA rule, denote $\mathcal{M}(j)$ as the set of channels to be sensed at Slot j, and if there are at least one sensed-idle channel, denote $K(j)$ as the channel to be accessed. So the reward of the rule is given as $U(t) = \sum_{j=1}^{t} \{S_{K(j)}(j) \cdot \mathbb{I}[\max_{i \in \mathcal{M}(j)} X_i(j) > 0]\}$, where $X_i(j)$ is the sensing observation of Channel i, and $\mathbb{I}[\max_{i \in \mathcal{M}(j)} X_i(j) > 0]$ indicates whether there are at least one sensed-idle channel. And the learning loss until Slot t is $R(t) = t \cdot \mathbb{E}[\max_{i \in \mathcal{M}^*} \mathbb{E}[S_i|X_i = 1]] - \mathbb{E}[U(t)]$. It is shown in [101] that the learning loss when sensing is perfect and the secondary user can access all sensed-idle channels is at least $O(\ln t)$ as $t \to \infty$.

5.2 Partial Channel Sensing and Access

In this section, we design our OCSA rule for single channel access with imperfect sensing. The first question is: what is the design challenge with the Markov model? In the i.i.d. model, for any two slots with whatever interval between them, the observations of a channel at the two slots are always independent and identically distributed. Therefore, based on observations of a channel at "scattered" slots, we can estimate the channel statistics (i.e., the channel idle probability), as we have done in Algorithm 4.3. However, with the Markov model, with randomly "scattered" observations, it is hard to accurately estimate the channel statistics [102]. To tackle the problem, inspired by the proof of Lemma 2.1 in [101], if a channel is selected to sense, we keep sensing it until two idle states are observed. So for each channel, we have scattered variable-length sensing periods, called *blocks*, and in each block, we have three phases: the first phase includes the first sampling until the sampling immediately prior to the first idle sampling; the second phase includes the first idle sampling until the sampling immediately prior to the second idle sampling; and the third phase includes the second idle sampling. If we concatenate the second phases of all the scattered blocks of a channel, say Channel i, we can construct a Markov chain. Based on the *strong Markov property*, the newly constructed Markov chain has the same statistics as the original Markov chain of Channel i. Then it suffices that we estimate the statistics of the newly constructed Markov chain. Accordingly we have a OCSA rule given in Algorithm 5.1, which is inspired by RCA-M in [102].

Remark 5.1. In the algorithm, the secondary user records T_i as the number of slots in

Algorithm 5.1 Single Channel Access with Partial Channel Sensing (Markov Model)

Initialization: $\mathcal{A} = \emptyset$ (null set), $\mathbf{I}_{ini} = \mathbf{0}$, $\mathbf{I}_{P2} = \mathbf{0}$, $t_2 = 0$, $Y_i = 0$, $T_i = 0$, $i \in \mathcal{N} = \{1, 2, ..., N\}$.

for each subsequent Slot t **do**

  if $|\mathcal{A}| < M$ **then**

    for $i := 1$ to N **do**

      if $\mathbf{I}_{ini}(i) = 0$ **then**

        $\mathcal{A} = \mathcal{A} \cup \{i\}$, $I_{ini}(i) = 1$.

    if $|\mathcal{A}| < M$ **then**

      Calculate an index of any channel $i \in \mathcal{N} \backslash \mathcal{A}$ as $\frac{Y_i}{T_i} + \sqrt{\frac{L_0 \ln t_2}{T_i}}$; from channel set $\mathcal{N} \backslash \mathcal{A}$, choose $(M - |\mathcal{A}|)$ channels with the largest indices, and then include these channels to $\mathcal{A}$.

  Sensing channels in $\mathcal{A}$, denote the set of channels sensed idle as $\mathcal{I}(t)$.

  for all channels $i \in \mathcal{A}$ **do**

    if $I_{P2}(i) = 0$ **then**

      if $i \in \mathcal{I}(t)$ **then**

        $I_{P2}(i) = 1, Y_i = Y_i + 1, T_i = T_i + 1$.

      else

        Do nothing

    else

      if $i \in \mathcal{I}(t)$ **then**

        $I_{P2}(i) = 0, \mathcal{A} = \mathcal{A} \backslash \{i\}$.

      else

        $T_i = T_i + 1$.

  if $\mathbf{I}_{P2} \neq \mathbf{0}$ **then**

    $t_2 = t_2 + 1$.

  if $\mathcal{I}(t) = \emptyset$ **then**

    Wait until slot $t + 1$.

  else

    Calculate indices $\frac{Y_i}{T_i} + \sqrt{\frac{L_0 \ln t_2}{T_i}}$, $i \in \mathcal{I}(t)$, access the channel with largest index.

phases 2 of Channel i's blocks, and Y_i as the number of slots in which Channel i is sensed idle in phases 2 of its blocks (in other words, Y_i is also the number of blocks of Channel i). So $\frac{Y_i}{T_i}$ is the sample mean of the probability of Channel i being idle in the newly constructed Markov chain. N-length vector $\mathbf{I}_{ini}$ indicates whether the channels have been sensed yet, with its ith element, denoted $I_{ini}(i)$, equal to 1 if Channel i has been sensed at least once, or equal to 0 otherwise. N-length vector $\mathbf{I}_{P2}$ indicates whether the channels are in phase 2 of their blocks, with its ith element, denoted $I_{P2}(i)$, equal to 1 if Channel i is in phase 2 of one of its blocks, or equal to 0 otherwise. Set $\mathcal{A}$ is the set of channels that are to be sensed. t_2 records the number of time slots in which there is at least one channel that is in phase 2 of one of its blocks. In the algorithm, Lines 2–8 are to determine $\mathcal{A}$, in which channels that are sensed in the previous slot have the highest priority (if the channels do not finish their blocks in the previous slot), channels that have not been sensed yet have the second highest priority, and other channels have the lowest priority. When those "other channels" are selected, similar to Algorithms 4.3 and 4.4, an extra term, which guarantees that each channel is sufficiently sensed, is added to the sample mean $\frac{Y_i}{T_i}$ to form the index. Lines 10–20 of Algorithm 5.1 are to update $\mathbf{I}_{P2}$, Y_i, and T_i ($i = 1, 2, ..., N$).

Theorem 5.1. When $L_0 \geq \frac{90}{\min\limits_{i=1,2,...,N}(2-p_{00}^i-p_{11}^i)}$, the regret $R(t)$ of Algorithm 5.1 is $O(\ln t)$ with $t \to \infty$ and with finite t.

Proof. For Channel i following Markov model with $\theta_i = \frac{1-p_{00}^i}{2-p_{00}^i-p_{11}^i}$ as the stationary idle probability, if it is sensed in successive time slots, and we denote the observations as $X_i(1), X_i(2), ...$, then for any $\delta_0 \geq 0$, we have [103]

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n} X_i(j) - f(\theta_i) \geq \delta_0\right) &\leq \frac{1+\frac{2-p_{00}^i-p_{11}^i}{10}\delta_0}{\left(\min\{f(\theta_i),1-f(\theta_i)\}\right)^{-\frac{1}{2}}} e^{-\delta_0^2 n \frac{2-p_{00}^i-p_{11}^i}{20}}, \\
\mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n} X_i(j) - f(\theta_i) \leq -\delta_0\right) &\leq \frac{1+\frac{2-p_{00}^i-p_{11}^i}{10}\delta_0}{\left(\min\{f(\theta_i),1-f(\theta_i)\}\right)^{-\frac{1}{2}}} e^{-\delta_0^2 n \frac{2-p_{00}^i-p_{11}^i}{20}}.
\end{aligned} \tag{5.1}$$

These two inequalities will be used in the following proof.

Without loss of generality, assume $\theta_1 \geq \theta_2 \geq ... \geq \theta_N$. Therefore, for the genie-aided case, the secondary user should always sense channels in set $\mathcal{M}^* = \{1, 2, ..., M\}$, and access the sensed-idle channel with the largest θ_i.

For our proposed OCSA rule, regret comes from sensing a different channel set from $\mathcal{M}^*$, or sensing channels in $\mathcal{M}^*$ but accessing a wrong channel.

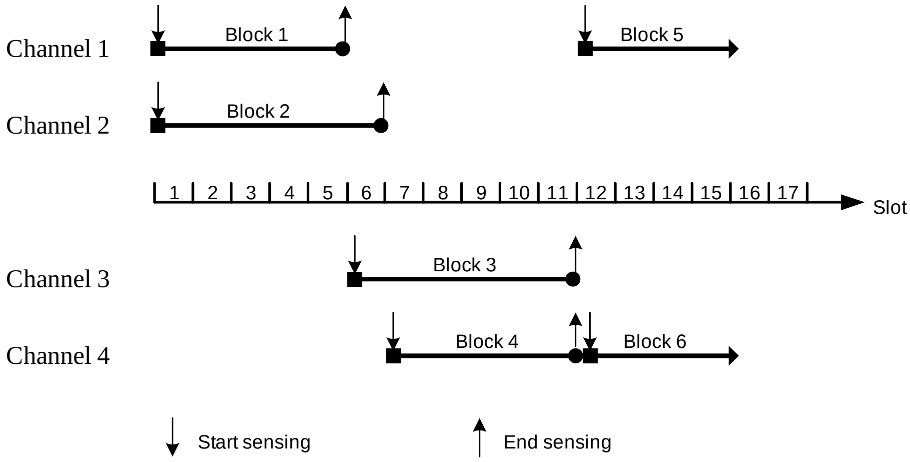

Fig. 5.2. An example of the blocks.

First consider that the secondary user senses a different channel set from $\mathcal{M}^*$. In specific, for a channel $i \notin \mathcal{M}^*$, we investigate the number of slots that Channel i is sensed. Recall that in our algorithm, if a channel is selected to sense, the channel will be sensed in successive slots until the second idle state of the channel is observed, and then a new channel (which may or may not be Channel i) will be decided by the secondary user to be added into the sensing set. Recall that a block is the sensing period of a channel until the second idle state is observed. Then a block is associated with a channel: at the beginning of the block, the secondary user starts sensing the channel until the end of the block. In this work, each block is assigned an index, starting from 1. Fig. 5.2 shows an example when the secondary user can sense two out of four channels at a slot. Channels 1 and 2 are sensed at Slot 1. So Block 1 (associated with Channel 1) and Block 2 (associated with Channel 2) start at Slot 1. At Slot 5, the second idle state of Channel 1 is observed, and thus, Block 1 ends at Slots 5. And a new channel, Channel 3, is added into the sensing set. Thus, Block 3 (associated with Channel 3) starts at Slot 6. At Slot 6, the second idle state of Channel 2 is observed, and thus, Block 2 ends at Slot 6. And a new channel, Channel 4, is added into the sensing set. Thus, Block 4 (associated with Channel 4) starts at Slot 7. At Slot 11, both Channels 3 and 4 have the second idle states, and thus, Blocks 3 and 4 end at Slot 11. And two channels, Channels 1 and 4 are decided by our proposed algorithm to sense from Slot 12. So Block 5 (associated with Channel 1) and Block 6 (associated with Channel 4) start at Slot 12. It can be seen that the blocks may overlap with each other.

We use $x^{(n)}$ to denote the value of x at the end of Block n, and $x(t)$ to denote the value of x at Slot t. For a worse channel $i \notin \mathcal{M}^*$, let b_i denote the number of completed blocks Channel i has been associated with. For any positive integer l, we have

$$
\begin{aligned}
b_i^{(n)} &= 1 + \sum_{j=N+1}^{n} \mathbb{I}\,[\text{Channel } i \text{ is associated with Block } j\,] \\
&\le l + \sum_{k=1}^{M}\sum_{j=N+1}^{n} \mathbb{I}\left[\frac{Y_k^{(j-1)}}{T_k^{(j-1)}} + \sqrt{\frac{L_0 \ln t_2^{(j-1)}}{T_k^{(j-1)}}} \le \frac{Y_i^{(j-1)}}{T_i^{(j-1)}} + \sqrt{\frac{L_0 \ln t_2^{(j-1)}}{T_i^{(j-1)}}}, b_i^{(j-1)} \ge l\right] \\
&\le l + \sum_{k=1}^{M}\sum_{j=N+1}^{n} \mathbb{I}\Big[\min_{0<s_1\le t_2^{(j-1)}} \frac{Y_k|_{T_k=s_1}}{s_1} + \sqrt{\frac{L_0 \ln t_2^{(j-1)}}{s_1}} \le \max_{l\le s_2\le t_2^{(j-1)}} \frac{Y_i|_{T_i=s_2}}{s_2} + \sqrt{\frac{L_0 \ln t_2^{(j-1)}}{s_2}}\Big] \\
&\le l + \sum_{k=1}^{M}\sum_{j=N+1}^{n}\sum_{s_1=1}^{t_2^{(j-1)}}\sum_{s_2=l}^{t_2^{(j-1)}} \mathbb{I}\left[\frac{Y_k|_{T_k=s_1}}{s_1} + \sqrt{\frac{L_0 \ln t_2^{(j-1)}}{s_1}} \le \frac{Y_i|_{T_i=s_2}}{s_2} + \sqrt{\frac{L_0 \ln t_2^{(j-1)}}{s_2}}\right] \\
&\le l + \sum_{k=1}^{M}\sum_{t_1=1}^{t_2^{(n)}}\sum_{s_1=1}^{t_1}\sum_{s_2=l}^{t_1} \mathbb{I}\left[\frac{Y_k|_{T_k=s_1}}{s_1} + \sqrt{\frac{L_0 \ln t_1}{s_1}} \le \frac{Y_i|_{T_i=s_2}}{s_2} + \sqrt{\frac{L_0 \ln t_1}{s_2}}\right]. \qquad (5.2)
\end{aligned}
$$

At Slot t, the number of completed blocks should be no more than t. So from (5.2) with $l = \left\lceil \frac{4L_0 \ln t}{\Delta_i} \right\rceil$ and $\Delta_i = (f(\theta_i) - f(\theta_M))^2$, we have the expected number of completed blocks associated with Channel i bounded as

$$
\begin{aligned}
\mathbb{E}[b_i(t)] \le& \left\lceil \frac{4L_0 \ln t}{\Delta_i} \right\rceil + \sum_{k=1}^{M}\sum_{t_1=1}^{t}\sum_{s_1=1}^{t_1}\sum_{s_2=\left\lceil \frac{4L_0 \ln t}{\Delta_i}\right\rceil}^{t_1} \\
&\left\{\mathbb{P}\left(\frac{Y_k|_{T_k=s_1}}{s_1} \le f(\theta_k) - \sqrt{\frac{L_0 \ln t_1}{s_1}}\right) + \mathbb{P}\left(\frac{Y_i|_{T_i=s_2}}{s_2} \ge f(\theta_i) + \sqrt{\frac{L_0 \ln t_1}{s_2}}\right)\right\} \\
\le& \left\lceil \frac{4L_0 \ln t}{\Delta_i} \right\rceil + \sum_{k=1}^{M}\sum_{t_1=1}^{t}\sum_{s_1=1}^{t_1}\sum_{s_2=\left\lceil \frac{4L_0 \ln t}{\Delta_i}\right\rceil}^{t_1} \left(\beta_i t_1^{-\frac{L_0(2-p_{00}^i-p_{11}^i)-10}{20}} + \beta_k t_1^{-\frac{L_0(2-p_{00}^k-p_{11}^k)-10}{20}}\right)
\end{aligned}
\qquad (5.3)
$$

where $\beta_i = \frac{2}{\left(\min\{f(\theta_i), 1-f(\theta_i)\}\right)^{-\frac{1}{2}}}$. Here the second line of (5.3) is similar to (4.18), and the third line comes from (5.1) and the fact that if we concatenate the observations in phases 2 of blocks of a channel, we form a new Markov chain with the same statistics as that of the original Markov chain that models the busy/idle states of the channel.

By setting $L_0 \ge 90/\min_{i=1,2,\ldots,N}(2 - p_{00}^i - p_{11}^i)$, we have an inequality

$$
\mathbb{E}\left[b_i(t)\right] \le \left\lceil \frac{4L_0 \ln t}{\Delta_i} \right\rceil + \sum_{k=1}^{M}\sum_{t_1=1}^{\infty}\sum_{s_1=1}^{t_1}\sum_{s_2=1}^{t_1} (\beta_i + \beta_k) {t_1}^{-4}
$$

$$\leq \frac{4L_0 \ln t}{\Delta_i} + \beta_i M \frac{\pi^2}{3} + \sum_{k=1}^{M} \beta_k \frac{\pi^2}{3} + 1. \tag{5.4}$$

By strong Markov property which guarantees the independence of the blocks associated with Channel i, and by Wald's lemma, the expected number of slots in which Channel i is sensed in completed blocks until Slot t is bounded by $\gamma_i \mathbb{E}[b_i(t)]$, in which $\gamma_i = 1 + \frac{2}{f(\theta_i)}$ is the expected length of an interval from the moment when a first idle state of Channel i is observed to the moment when the third idle state of Channel i is observed, and is finite. Here γ_i is an upper bound of the expected block length of Channel i. Then, the average number of slots where a non-optimal set is sensed in completed blocks until Slot t is bounded by

$$\mathbb{E}[T_{\ddagger}(t)] \leq \sum_{i=M+1}^{N} \gamma_i \mathbb{E}[b_i(t)] \leq \ln t \sum_{i=M+1}^{N} \frac{4L_0\gamma_i}{\Delta_i} + \sum_{i=M+1}^{N} \left(\gamma_i \beta_i M \frac{\pi^2}{3} + \sum_{k=1}^{M} \gamma_i \beta_k \frac{\pi^2}{3} + \gamma_i \right) \tag{5.5}$$

where $\ddagger$ means non-optimality.

Next we consider that the secondary user senses the optimal channel set $\mathcal{M}^*$ but accesses a wrong (suboptimal) channel. Let $T_{*,\ddagger}(t)$ denote number of slots in completed blocks until Slot t where $\mathcal{M}^*$ is sensed but a wrong channel is accessed. Here the first subscript $*$ means optimal sensing set, and the second subscript $\ddagger$ means wrong channel access. Let $T_{*,i^\ddagger}(t)$ denote the number of slots in completed blocks until Slot t where $\mathcal{M}^*$ is sensed and Channel i is wrongly accessed. So we have $T_{*,\ddagger}(t) = \sum_{i=2}^{M} T_{*,i^\ddagger}(t)$.[1] Here we use $x^{[n]}$ to denote the value of x at the slot when the optimal sensing set $\mathcal{M}^*$ is sensed the nth time slot. When $\mathcal{M}^*$ is sensed the jth time, let $\mathcal{I}_{\mathcal{M}*}^{[j]}$ denote the set of sensed-idle channels, and $k^{[j]}$ denote the channel selected to access. So the number of slots in which $\mathcal{M}^*$ is sensed but channel i is wrongly accessed until the slot when $\mathcal{M}^*$ is sensed the nth time is given as

$$\begin{aligned}
T_{*,i^\ddagger}^{[n]} &= \sum_{j=1}^{n} \mathbb{I}\left[k^{[j]} = i, i > \min_{m \in \mathcal{I}_{\mathcal{M}*}^{[j]}} m\right] \\
&\leq l + \sum_{j=1}^{n} \sum_{\mathcal{S} \subset \mathcal{M}^*} \mathbb{I}\left[k^{[j]} = i, \mathcal{I}_{\mathcal{M}*}^{[j]} = \mathcal{S}, i > \min_{m \in \mathcal{S}} m, T_{*,i^\ddagger}^{[j-1]} \geq l\right] \\
&\leq l + \sum_{j=1}^{n} \sum_{k=1}^{i-1} \mathbb{I}\left[\frac{Y_k^{[j]}}{T_k^{[j]}} + \sqrt{\frac{L_0 \ln t_2^{[j]}}{T_k^{[j]}}} \leq \frac{Y_i^{[j]}}{T_i^{[j]}} + \sqrt{\frac{L_0 \ln t_2^{[j]}}{T_i^{[j]}}}, T_{*,i^\ddagger}^{[j-1]} \geq l\right]
\end{aligned}$$

[1]Since Channel 1 has the largest idle probability, it is always optimal to access Channel 1 as long as Channel 1 is sensed idle.

$$\leq l + \sum_{k=1}^{i-1}\sum_{j=1}^{n}\mathbb{I}\left[\min_{0<s_1\leq t_2^{[j]}}\frac{Y_k|_{T_k=s_1}}{s_1}+\sqrt{\frac{L_0\ln t_2^{[j]}}{s_1}}\leq \max_{l\leq s_2\leq t_2^{[j]}}\frac{Y_i|_{T_i=s_2}}{s_2}+\sqrt{\frac{L_0\ln t_2^{[j]}}{s_2}}\right]$$

$$\leq l + \sum_{k=1}^{i-1}\sum_{j=1}^{n}\sum_{s_1=1}^{t_2^{[j]}}\sum_{s_2=l}^{t_2^{[j]}}\mathbb{I}\left[\frac{Y_k|_{T_k=s_1}}{s_1}+\sqrt{\frac{L_0\ln t_2^{[j]}}{s_1}}\leq \frac{Y_i|_{T_i=s_2}}{s_2}+\sqrt{\frac{L_0\ln t_2^{[j]}}{s_2}}\right]$$

$$\leq l + \sum_{k=1}^{i-1}\sum_{t_1=1}^{t_2^{[n]}}\sum_{s_1=1}^{t_1}\sum_{s_2=l}^{t_1}\mathbb{I}\left[\frac{Y_k|_{T_k=s_1}}{s_1}+\sqrt{\frac{L_0\ln t_1}{s_1}}\leq \frac{Y_i|_{T_i=s_2}}{s_2}+\sqrt{\frac{L_0\ln t_1}{s_2}}\right].$$

Then, similar to inequalities (5.3) and (5.4), we can bound expectation of $T_{*,i^\ddagger}(t)$ as

$$\begin{aligned}\mathbb{E}[T_{*,i^\ddagger}(t)] &= \mathbb{E}\left[\sum_{n=1}^{t}\mathbb{I}\left[\mathcal{M}^* \text{ is sensed } n \text{ times until slot } t\right]T_{*,i^\ddagger}^{[n]}\right]\\ &\leq l+\sum_{k=1}^{i-1}\sum_{t_1=1}^{t}\sum_{s_1=1}^{t_1}\sum_{s_2=l}^{t_1}\mathbb{E}\left[\mathbb{I}\left[\frac{Y_k|_{T_k=s_1}}{s_1}+\sqrt{\frac{L_0\ln t_1}{s_1}}\leq\frac{Y_i|_{T_i=s_2}}{s_2}+\sqrt{\frac{L_0\ln t_1}{s_2}}\right]\right]\\ &\leq \frac{4L_0\ln t}{\Delta_{i,i-1}}+(i-1)\beta_i\frac{\pi^2}{3}+\sum_{k=1}^{i-1}\beta_k\frac{\pi^2}{3}+1\end{aligned} \tag{5.6}$$

where $\Delta_{i,k} = (f(\theta_i) - f(\theta_k))^2$.

Therefore, we have

$$\begin{aligned}\mathbb{E}\left[T_{*,\ddagger}(t)\right] &= \sum_{i=2}^{M}\mathbb{E}\left[T_{*,i^\ddagger}(t)\right]\\ &\leq \ln t\sum_{i=2}^{M}\frac{4L_0}{\Delta_{i,i-1}}+\sum_{i=2}^{M}(i-1)\beta_i\frac{\pi^2}{3}+\sum_{i=2}^{M}\sum_{k=1}^{i-1}\beta_k\frac{\pi^2}{3}+M-1.\end{aligned} \tag{5.7}$$

In the following, the regret $R(t)$ is calculated. We denote the slot index of the last completed block until Slot t as $T(t)$. We use $r_{*,*}$ to denote the expected reward at a slot by sensing $\mathcal{M}^*$ and accessing the optimal channel.

$$\begin{aligned}R(t) &= t\cdot r_{*,*} - \mathbb{E}\Big[\sum_{j=1}^{t}\mathbb{I}\big[\max_{i\in\mathcal{M}(j)}X_i(j)=1\big]S_{k(j)}(j)\Big]\\ &= \big\{\mathbb{E}\left[T(t)\right]\cdot r_{*,*} - \mathbb{E}[T_{*,*}(t)]\cdot r_{*,*}\big\}\\ &+\Big\{\mathbb{E}[T_{*,*}(t)]\cdot r_{*,*}-\mathbb{E}\Big[\sum_{j=1}^{T(t)}\mathbb{I}\big[\max_{i\in\mathcal{M}(j)}X_i(j)=1\big]S_{k(j)}(j)\Big]\Big\}\\ &+\Big\{\mathbb{E}[t-T(t)]\cdot r_{*,*}-\mathbb{E}\Big[\sum_{j=T(t)+1}^{t}\mathbb{I}\big[\max_{i\in\mathcal{M}(j)}X_i(j)=1\big]S_{k(j)}(j)\Big]\Big\}.\end{aligned} \tag{5.8}$$

It can be seen that the regret $R(t)$ is decomposed into three parts. In the following we analyze them one by one.

The first part in (5.8) is

$$\mathbb{E}\left[T(t)\right] \cdot r_{*,*} - \mathbb{E}\left[T_{*,*}(t)\right] \cdot r_{*,*} = r_{*,*}(\mathbb{E}\left[T_{\ddagger}(t)\right] + \mathbb{E}\left[T_{*,\ddagger}(t)\right]) \tag{5.9}$$

which is bounded by $O(\ln t)$ from (5.5) and (5.7).

Now we investigate the second part in (5.8). For our OCSA algorithm, consider the channel sensing and access states at successive slots. Denote $\bar{b}(t)$ as the number of transitions from “sense $\mathcal{M}^*$ and access the best idle channel" to any other states (“not sense $\mathcal{M}^*$" or “sense $\mathcal{M}^*$ but not access the best idle channel") until $T(t)$. Therefore, from Slot 1 to Slot $T(t)$, we have $\bar{b}(t) + 1$ scattered *intervals*: in each interval the secondary user senses $\mathcal{M}^*$ and accesses the best idle channel. For the vth interval, we divide it to two partitions: the first partition is until the slot by which all channels have been sensed idle at least once, and the second partition is the rest in the interval. Note that it is possible that the second partition does not exist in some intervals. Denote the lengths of the first and the second partition as $x_1^{\{v\}}$ and $x_2^{\{v\}}$, respectively, and denote the total reward in the first and second partition as $r_1^{\{v\}}$ and $r_2^{\{v\}}$, respectively. Then the second partitions of the $\bar{b}(t) + 1$ intervals are independent from each other.

Then, we bound the second part is (5.8) as

$$\begin{aligned}
&\mathbb{E}\left[T_{*,*}(t)\right] \cdot r_{*,*} - \mathbb{E}\left[\sum_{j=1}^{T(t)} \mathbb{I}\left[\max_{i\in\mathcal{M}(j)} X_i(j) = 1\right] S_{k(j)}(j)\right] \\
&\leq \mathbb{E}\left[T_{*,*}(t)\right] \cdot r_{*,*} - \mathbb{E}\left[\sum_{v=1}^{\bar{b}(t)+1} (r_1^{\{v\}} + r_2^{\{v\}})\right] \\
&= r_{*,*}\mathbb{E}\left[\sum_{v=1}^{\bar{b}(t)+1} \left(x_1^{\{v\}} + x_2^{\{v\}}\right)\right] - \mathbb{E}\left[\sum_{v=1}^{\bar{b}(t)+1} (r_1^{\{v\}} + r_2^{\{v\}})\right] \\
&\leq r_{*,*}\mathbb{E}\left[\sum_{v=1}^{\bar{b}(t)+1} x_1^{\{v\}}\right] + r_{*,*}\mathbb{E}\left[\sum_{v=1}^{\bar{b}(t)+1} x_2^{\{v\}}\right] - \mathbb{E}\left[\sum_{v=1}^{\bar{b}(t)+1} r_2^{\{v\}}\right] \\
&\leq r_{*,*}\mathbb{E}\left[\bar{b}(t) + 1\right] \sum_{i=1}^{M} \left(\frac{1}{f(\theta_i)} + 1\right).
\end{aligned} \tag{5.10}$$

TABLE 5.1
PARAMETERS USED IN THE SIMULATION.

$N = 5$	θ_i (stationary)	(0.5296,0.4001,0.9817,0.1931,0.2495)
	P^i_{00}	(0.7914,0.9372,0.3080,0.9793,0.8778)
	P^i_{11}	(0.8147,0.9058,0.9871,0.9134,0.6324)
$N = 6$	θ_i (stationary)	(0.3605,0.9291,0.7694,0.6199,0.4109,0.3559)
	P^i_{00}	(0.7450,0.2806,0.7077,0.0459,0.4178,0.9242)
	P^i_{11}	(0.5476,0.9451,0.9124,0.4150,0.1653,0.8628)
$N = 7$	θ_i (stationary)	(0.8811,0.5390,0.3468,0.9522,0.7823,0.0471,0.7968)
	P^i_{00}	(0.9044,0.1652,0.7675,0.2490,0.2838,0.9550,0.6355)
	P^i_{11}	(0.9871,0.2860,0.5621,0.9623,0.8007,0.0900,0.9071)
$N = 8$	θ_i (stationary)	(0.6923,0.5430,0.3544,0.8753,0.5212,0.6759,0.8783,0.9762)
	P^i_{00}	(0.3486,0.4313,0.7378,0.2363,0.7159,0.3479,0.1595,0.2371)
	P^i_{11}	(0.7105,0.5214,0.5224,0.8912,0.7390,0.6873,0.8835,0.9814)

Notice that $\mathbb{E}\left[\bar{b}(t)\right] \leq \sum_{i=M+1}^{N} \mathbb{E}\left[b_i(t)\right] + \mathbb{E}\left[T_{*,\ddagger}(t)\right]$. From (5.4) and (5.7), it can be seen that the second part in (5.8) is bounded by $O(\ln t)$.

The third part in (5.8) is upper bounded as

$$\mathbb{E}\left[t - T(t)\right] r_{*,*} - \mathbb{E}\left[\sum_{j=T(t)+1}^{t} \mathbb{I}\left[\max_{i \in \mathcal{M}(j)} X_i(j) = 1\right] S_{k(j)}(j)\right]$$
$$\leq r_{*,*}\mathbb{E}\left[t - T(t)\right] \leq r_{*,*} \max_{i \in \{1,2,\ldots,N\}} \frac{1}{f(\theta_i)}. \tag{5.11}$$

By (5.8), (5.9), (5.10) and (5.11), we have our regret $R(t)$ bounded by $O(\ln t)$. □

5.3 Performance Evaluation

We use Monte-Carlo simulation to validate our analysis. Consider a cognitive radio network with $N = 5, 6, 7, 8$ channels. Homogeneous sensing is considered with $P^i_d = 0.8$ and $P^i_f = 0.3$ for any i. and other parameters are listed in Table 5.1.

Our proposed algorithm for the Markov model is evaluated, and Fig. 5.3 shows the average $R(t)/\ln t$ in homogeneous sensing and partial channel sensing. It can be seen that the normalized regrets are finitely bounded, which is consistent with our analysis.

The impact of K in homogeneous sensing and partial channel sensing with the Markov model is shown in Fig. 5.4. It can be seen that, for homogeneous sensing and partial channel

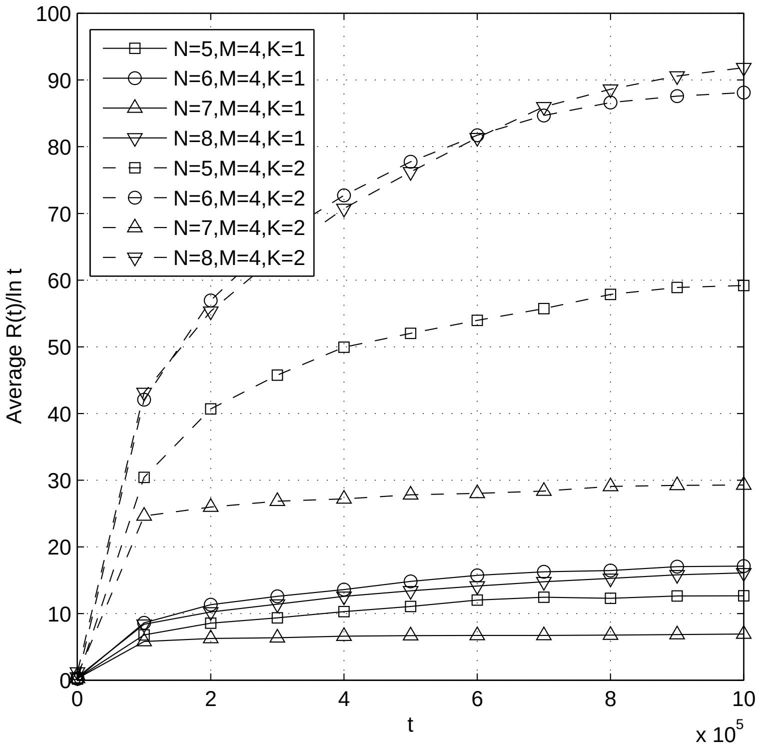

Fig. 5.3. Average $R(t)/\ln t$ of the proposed algorithm for Markov model with homogeneous sensing and partial channel sensing.

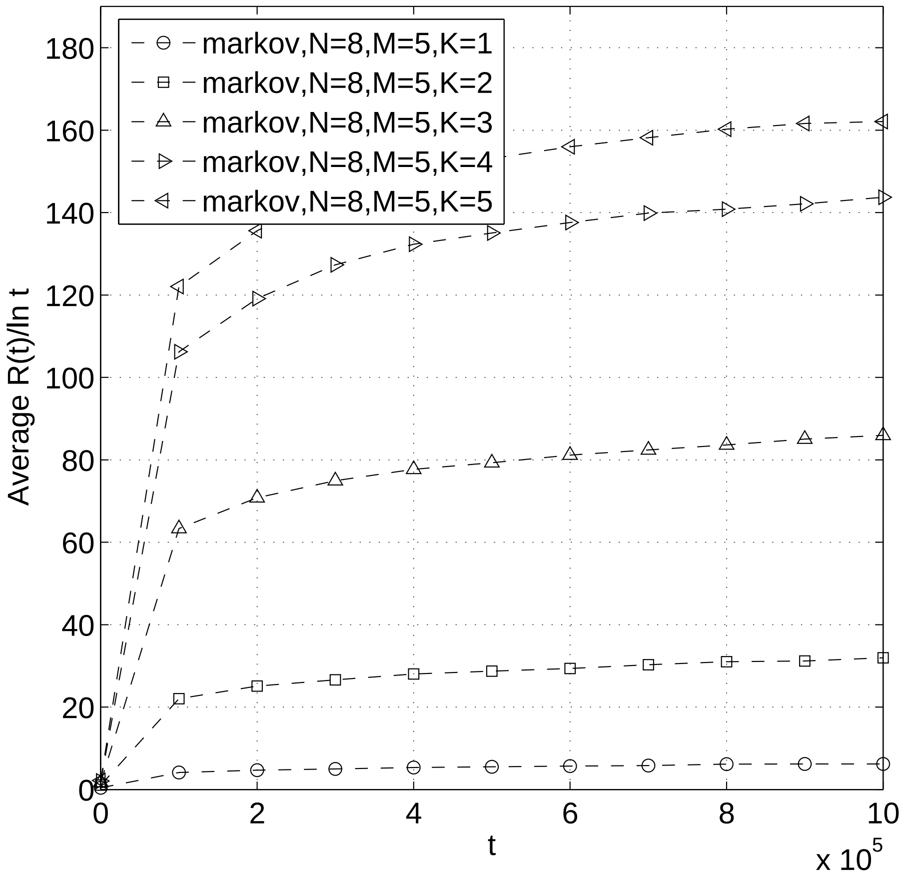

Fig. 5.4. Impact of K on the average $R(t)/\ln t$ of the proposed algorithms in homogeneous sensing and partial channel sensing with the Markov model.

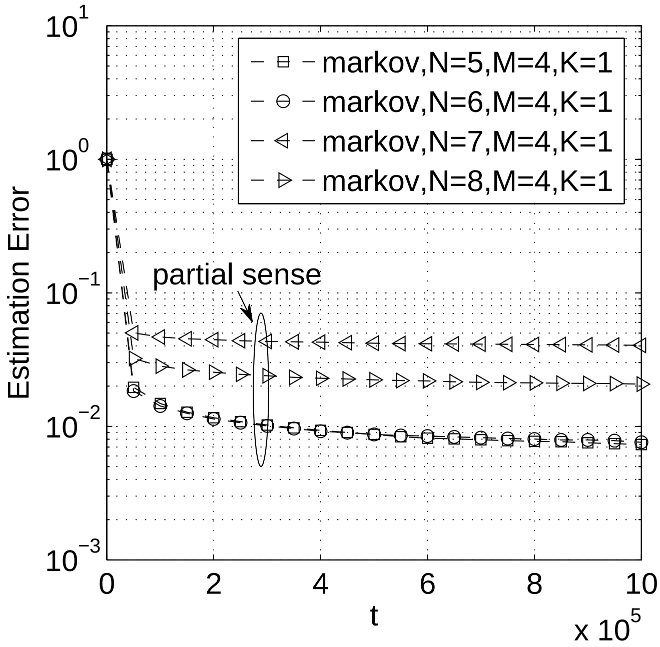

Fig. 5.5. Estimation errors of $\mathbf{\Theta} = (\theta_1, \theta_2, ..., \theta_N)$, Markov model.

sensing with the Markov model, the normalized regret increases when K increases from 1 to M. This is because, in partial channel sensing, the M sensed channels in our algorithms and in the genie-aided rule may not be the same. If we select more channels (larger K) to access, then the channels selected by our algorithms may have bigger difference from the channels selected by the genie-aided rule, and thus, the normalized regret increases.

Then, we evaluate the accuracy of the estimators for the stationary channel idle probability θ_i's in the Markov model. Fig. 5.5 shows the estimation errors for the Markov model in homogeneous sensing and partial channel sensing (denoted as "markov"). Here the estimation error is defined as ratio of L_2-norm of $(\mathbf{\Theta} - \hat{\mathbf{\Theta}})$ to L_2-norm of $\mathbf{\Theta}$ (i.e., $\frac{||\mathbf{\Theta}-\hat{\mathbf{\Theta}}||_2}{||\mathbf{\Theta}||_2}$) where $\mathbf{\Theta}$ and $\hat{\mathbf{\Theta}}$ are the vector of real channel idle probabilities and the vector of estimated channel idle probabilities, respectively. It can be seen that the estimation error tends to converge to 0 when t keeps increasing. This can be explained as follows. For partial channel sensing in the Markov model, as shown at Algorithm 5.1, the extra term in the indices determines that when t increases, the secondary user mainly senses the best channel (the channel with the largest θ_i), while other inferior channels will still be sensed in the scale of $O(\ln t)$. Therefore, the estimation error should also converge to 0 when $t \to \infty$, although the convergence speed is much smaller than that in the full channel sensing case in Fig. 4.9 of Chapter 4.

5.4 Conclusion

In this chapter, the problem of OCSA in a CRN is investigated, where the Markov model of channel idle/busy states over time slots is assumed. In the case with partial channel sensing, an OCSA rule is proposed, and under mild condition (i.e., the condition of L_0 in Theorem 5.1) the regret $O(\ln t)$ can be achieved for a finite time t and $t \to \infty$. The performance of our proposed rule is verified through numerical computation. This research should provide insights to the design of OCSA in CRNs with unknown statistical information of primary traffic.

In the homogeneous sensing considered in this chapter, the stationary channel idle probabilities can be estimated by using samples of the newly constructed Markov chain. It may not be feasible to extend this idea to OCSA design with heterogeneous sensing. The reason is that the information of estimated stationary channel idle probabilities does not help in designing an order optimal algorithm for heterogeneous sensing (recalling that in Chapter 4, we use reward rather than channel idle probability as criterion in designing our order optimal algorithm for heterogeneous sensing). So the heterogeneous sensing with Markov model deserves future investigation.

Chapter 6

Optimal Distributed OCA in Wireless AF Relay Networks

In this chapter, we investigate distributed OCA in a cooperative network with multiple AF relays. Two cases are considered: *Case I with full CSI at a winner source* where a winner source in a contention has CSI of links from itself to relays and from relays to its destination, and *Case II with partial CSI at a winner source* where a winner source only has CSI of links from itself to relays. In Case I, it is found that a pure-threshold strategy exists to optimize the average system throughput. There are two stopping problems in Case II, one in the main layer (for channel access of sources) and the other in the sub-layer (for channel access of relays). An intuitive strategy is proposed, which is shown to be non-optimal. We also theoretically derive an optimal strategy for Case II. In either the intuitive strategy or the optimal strategy, the first-hop stopping rule has a pure-threshold structure, while the second-hop stopping rule has a threshold determined by channel gain realization in the preceding first-hop transmission. Numerical results are presented to demonstrate the effectiveness and efficiency of proposed strategies.[1]

6.1 Introduction

OCA has received much attention in the literature, particularly in centralized networks [27], [32]. A central controller can collect the CSI of the users, and schedule only those users

[1] *A version of this chapter has been published in IEEE Journal on Selected Areas in Communications, 30: 1675–1683 (2012).*

with the best channel conditions. On the other hand, the research on optimal distributed OCA is still in its infancy. As discussed in Section 2.1.3, existing research efforts for optimal distributed OCA are for single-hop peer-to-peer communications in an ad hoc network. As wireless relaying has recently attracted a lot of research interests [104]–[114], in this chapter we investigate optimal distributed OCA in a wireless AF relay network. We consider multiple source-destination pairs aided by multiple relays. Since transmission between each source-destination pair involves two hops: from source to relays and from relays to the destination, the problem of OCA in a relay network is quite different from those in a single-hop network (e.g., in reference [33]–[35]), and is challenging as multi-source diversity, multi-relay diversity, and time diversity should be all exploited.

6.2 Case I: with Full CSI at a Winner Source

6.2.1 System Model

Consider K source-destination pairs aided by L relays, as shown in Fig. 6.1. For transmission from a source to its destination, there is no direct link, and only one relay is selected to help with AF mode. The transmission power of a source and a relay is P_s and P_r, respectively. Channel reciprocity in terms of channel gain is assumed, and we denote the channel gain from the ith source to the jth relay (and vice versa) as f_{ij}, the channel gain from the jth relay to the ith destination (and vice versa) as g_{ji}. Assume f_{ij} and g_{ji} follow a complex Gaussian distribution with mean being zero and variance being σ_f^2 and σ_g^2, respectively. Noise is assumed to be Gaussian with unit variance. For source-to-destination transmission, say from the ith source to its destination aided by the jth relay, the maximal rate that can be achieved in AF mode is

$$\log_2\left(1+\frac{P_sP_r|f_{ij}|^2|g_{ji}|^2}{1+P_s|f_{ij}|^2+P_r|g_{ji}|^2}\right) \tag{6.1}$$

and the data transmission time from the source to the relay and from the relay to the destination are both $\frac{\tau_d}{2}$.

Channel contention of the sources is as follows. At the beginning of a time slot with duration δ,[2] each source independently contends for the channel by sending a request-to-

[2]Note that this time slot used in the channel contention, which usually has length of a few microseconds, is different from the time slot used in Chapters 3–5.

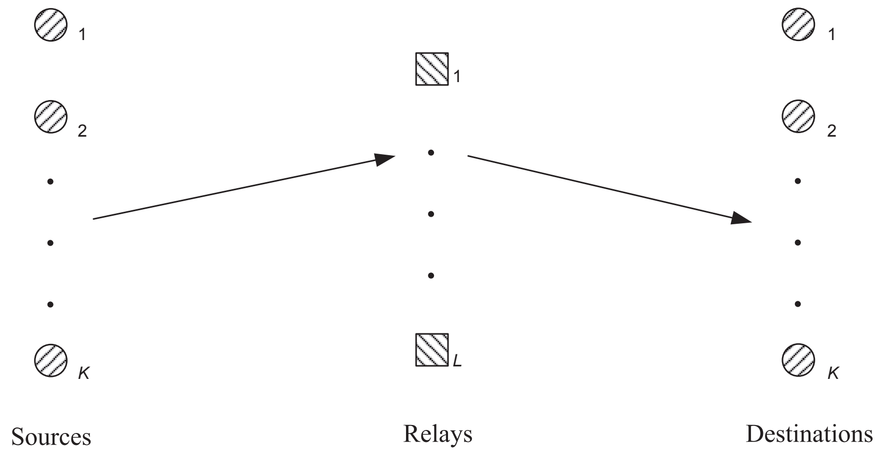

Fig. 6.1. System model.

send (RTS) packet[3] with probability p_0. There are three possible outcomes:

- If there is no source transmitting RTS in the time slot (with probability $(1-p_0)^K$), then all the sources continue to contend in the next time slot;
- If there are two or more sources transmitting RTS (with probability $1-(1-p_0)^K - Kp_0(1-p_0)^{K-1}$), a collision happens, and then in the next time slot after the RTS transmission all sources continue to contend;
- If there is only one source, say Source i, transmitting RTS (with probability $Kp_0(1-p_0)^{K-1}$), then Source i is called *winner* of the contention. By reception of the RTS, each relay can estimate CSI between Source i and itself. Then the first relay transmits an RTS to Destination i, and Destination i replies with a clear-to-send (CTS) packet[4], which can be received by all relays. By reception of the CTS from Destination i, each relay can estimate its CSI with Destination i. Then all relays send a CTS to Source i in turn. In the CTS from a relay to Source i, CSI of the relay with Source i and with Destination i is included. After reception of the CTSs, Source i knows CSI from itself to all relays and from all relays to its destination. Then Source i has two options: 1) Source i selects the relay that renders its maximal source-to-destination rate, i.e., Source i selects Relay $j^* = \arg \max_{j\in\{1,\ldots,L\}} \left\{\log_2\left(1+\frac{P_sP_r|f_{ij}(n)|^2|g_{ji}(n)|^2}{1+P_s|f_{ij}(n)|^2+P_r|g_{ji}(n)|^2}\right)\right\}$ and

[3]In this chapter, different from traditional RTS in IEEE 802.11, training signal is embedded, but the duration value (which specifies the total duration of handshake between the transmitter and receiver) is not included.

[4]In this chapter, different from traditional CTS in IEEE 802.11, channel gain information is embedded, but the duration value is not included.

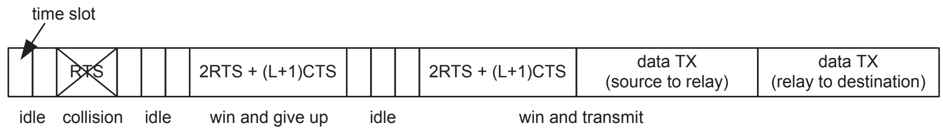

Fig. 6.2. An example of channel contention of sources.

transmits its packet to Relay j^* within duration $\frac{\tau_d}{2}$, then Relay j^* forwards the packet to Destination i within duration $\frac{\tau_d}{2}$; or 2) Source i gives up its transmission opportunity, and other sources can detect an idle slot after the RTS and CTS exchanges among Source i, all relays, and Destination i (i.e., that idle slot tells other sources that Source i gives up its transmission opportunity). After that a new contention is started among all the source nodes.

An example of the channel contention procedure is shown in Fig. 6.2. In the example, no source transmits RTS in the first two time slots. Then two or more sources transmit RTS, which results in a collision. After three idle slots, one winner appears. However, it gives up its transmission opportunity. Then after three more idle slots (the first is used to indicate the previous winner gives up, and the other two are for two new contentions), another winner appears. After exchange of 2 RTSs and $(L+1)$ CTSs, the winner transmits its data to its selected relay and the relay forwards the data to the winner's destination.

6.2.2 Optimal Stopping Strategy

Define an *observation* as the process of channel contention among the sources until a successful contention (i.e., a winner source appears). In an observation, the number of contentions follows a geometric distribution with parameter $Kp_0(1-p_0)^{K-1}$. Among all the contentions in an observation, the last contention is successful, with total duration (excluding data transmission) $2\tau_{RTS}+(L+1)\tau_{CTS}$ where τ_{RTS} and τ_{CTS} are duration of an RTS and CTS, respectively; and any other contention is either an idle slot (with duration δ) or a collision (with duration τ_{RTS}). The mean of the duration of an observation is then given as $\tau_o = 2\tau_{RTS}+(L+1)\tau_{CTS}+\frac{(1-p_0)^K}{Kp_0(1-p_0)^{K-1}}\cdot\delta+\frac{1-(1-p_0)^K-Kp_0(1-p_0)^{K-1}}{Kp_0(1-p_0)^{K-1}}\cdot\tau_{RTS}$.

After each observation, the winner source decides whether to continue a new observation (i.e., a new contention is started) or to stop (i.e., the winner source transmits its data). In the nth observation, let $s(n)$ denote the winner source. Then the observed information in the nth observation is $X(n) := \{s(n), f_{s(n)1}(n), ..., f_{s(n)L}(n), g_{1s(n)}(n), ..., g_{Ls(n)}(n)\}$. Here f and g with index (n) means the channel gain realizations at the end of the nth observation. For the nth observation, the reward Y_n is the total traffic volume that can be sent if the winner source transmits its data, which is a function of $X(n)$, and the cost T_n is the total waiting time from the first observation until the nth observation plus the data transmission time. If it is decided to stop at the Nth observation, then the average system throughput is $\frac{Y_N}{T_N}$. In the sequel, capital N is called the *stopping time*. And our objective is to find the optimal stopping time (also called optimal stopping strategy), N^*, which attains the maximal average system throughput $\sup\limits_{N\geq 0} \frac{\mathbb{E}[Y_N]}{\mathbb{E}[T_N]}$. According to [36, Chapter 6], this maximal-expected-return problem can be equivalently transformed into a standard form with its reward being $(Y_N - \lambda^* T_N)$. In particular, to get N^*, we need to find an optimal strategy to reach maximal expected reward

$$V^*(\lambda^*) = \sup_{N\geq 0} \{\mathbb{E}[Y_N] - \lambda^* \mathbb{E}[T_N]\} \tag{6.2}$$

where λ^* satisfies $\sup\limits_{N\geq 0} \{\mathbb{E}[Y_N] - \lambda^* \mathbb{E}[T_N]\} = 0$. Here λ^* is actually the maximal system throughput in our problem. This transformation method will be used when we solve the optimal stopping problems in our research, as shown in the sequel.

To formulate our research problem as an optimal stopping problem, in the nth observation, the reward is $Y_n = \frac{\tau_d}{2} R_n$ with the spent time denoted as $T_n = \sum\limits_{l=1}^{n} t_l + \tau_d$ where R_n is the achievable rate of the winner source in the nth observation via the best relay, given as

$$R_n = \sum_{i=1}^{K} \mathbb{I}[s(n) = i] \max_{j\in\{1,...,L\}} \left\{ \log_2 \left(1 + \frac{P_s P_r |f_{ij}(n)|^2 |g_{ji}(n)|^2}{1 + P_s |f_{ij}(n)|^2 + P_r |g_{ji}(n)|^2} \right) \right\}. \tag{6.3}$$

Here $\mathbb{I}[\cdot]$ means an indicator function, and t_l is the time spent in the lth observation with mean being τ_o. For finding a strategy N^* to achieve maximal average system throughput $\frac{\mathbb{E}[Y_N]}{\mathbb{E}[T_N]}$, it is equivalent [36] to design a strategy which attains

$$V^*(\lambda^*) = \sup_{N\geq 0} \left\{ \frac{\tau_d}{2} \mathbb{E}[R_N] - \lambda^* \mathbb{E} \left[\tau_d + \sum_{l=1}^{n} t_l \right] \right\} \tag{6.4}$$

where λ^* satisfies $V^*(\lambda^*) = 0$.

Before deriving an optimal stopping strategy N^*, two conditions

$$\mathbb{E}\Big[\sup_n\{Y_n - \lambda T_n\}\Big] = \mathbb{E}\Big[\sup_n\big\{\frac{\tau_d}{2}R_n - \lambda(\tau_d + \sum_{l=1}^{n} t_l)\big\}\Big] < \infty$$

and

$$\limsup_{n\to\infty}\big\{Y_n - \lambda T_n\big\} = \limsup_{n\to\infty}\big\{\frac{\tau_d}{2}R_n - \lambda\big(\tau_d + \sum_{l=1}^{n} t_l\big)\big\} = -\infty \; a.s.$$

should be checked which guarantee the existence of an optimal stopping strategy. Here λ can be viewed as the system throughput, while λ^* has the physical meaning of maximal system throughput.

Lemma 6.1. The first condition is satisfied as $\mathbb{E}\big[\sup_{n>0}\{\frac{\tau_d}{2}R_n - \lambda(\tau_d + \sum_{l=1}^{n} t_l)\}\big] < \infty$.

Proof. The mean of achievable transmission rate at the nth observation is

$$\mathbb{E}[R_n] = \sum_{i=1}^{K}\frac{1}{K}\mathbb{E}\left[\max_{j\in\{1,\ldots,L\}}\left\{\log_2\left(1+\frac{P_sP_r|f_{ij}(n)|^2|g_{ji}(n)|^2}{1+P_s|f_{ij}(n)|^2+P_r|g_{ji}(n)|^2}\right)\right\}\right]. \tag{6.5}$$

Since f_{ij} and g_{ji} follow complex Gaussian distribution with mean being zero and variance being σ_f^2 and σ_g^2, respectively, we have $\mathbb{E}[|f_{ij}|^2] = \sigma_f^2$ and $\mathbb{E}[|g_{ji}|^2] = \sigma_g^2$. Then we have

$$\begin{aligned}
\mathbb{E}[R_n] &= \sum_{i=1}^{K}\tfrac{1}{K}\mathbb{E}\left[\max_{j\in\{1,\ldots,L\}}\left\{\log_2\left(1+\tfrac{P_sP_r|f_{ij}(n)|^2|g_{ji}(n)|^2}{1+P_s|f_{ij}(n)|^2+P_r|g_{ji}(n)|^2}\right)\right\}\right]\\
&< \sum_{i=1}^{K}\tfrac{1}{K}\mathbb{E}\left[\textstyle\sum_{j=1}^{L}\log_2\left(1+\tfrac{P_sP_r|f_{ij}(n)|^2|g_{ji}(n)|^2}{1+P_s|f_{ij}(n)|^2+P_r|g_{ji}(n)|^2}\right)\right]\\
&\overset{(c)}{\le} \sum_{i=1}^{K}\tfrac{1}{K}\textstyle\sum_{j=1}^{L}\tfrac{1}{\ln 2}\mathbb{E}[P_s|f_{ij}|^2]\mathbb{E}[P_r|g_{ji}|^2]\\
&= \tfrac{1}{\ln 2}LP_sP_r\sigma_f^2\sigma_g^2 < \infty
\end{aligned} \tag{6.6}$$

$$\begin{aligned}
\mathbb{E}[R_n^2] &= \sum_{i=1}^{K}\tfrac{1}{K}\mathbb{E}\left[\left(\max_{j\in\{1,\ldots,L\}}\left\{\log_2\left(1+\tfrac{P_sP_r|f_{ij}(n)|^2|g_{ji}(n)|^2}{1+P_s|f_{ij}(n)|^2+P_r|g_{ji}(n)|^2}\right)\right\}\right)^2\right]\\
&< \sum_{i=1}^{K}\tfrac{1}{K}\mathbb{E}\left[\textstyle\sum_{j=1}^{L}\log_2^2\left(1+\tfrac{P_sP_r|f_{ij}(n)|^2|g_{ji}(n)|^2}{1+P_s|f_{ij}(n)|^2+P_r|g_{ji}(n)|^2}\right)\right]\\
&\overset{(d)}{\le} \sum_{i=1}^{K}\tfrac{1}{K}\textstyle\sum_{j=1}^{L}\tfrac{1}{(\ln 2)^2}\mathbb{E}[P_s^2|f_{ij}|^4]\mathbb{E}[P_r^2|g_{ji}|^4]\\
&= \tfrac{4}{(\ln 2)^2}LP_s^2P_r^2\sigma_f^4\sigma_g^4 < \infty
\end{aligned} \tag{6.7}$$

where (c) and (d) come from the fact that for $x, y \geq 0$, we have

$$\log_2 \Big(1 + \frac{xy}{1+x+y}\Big) \leq \frac{\frac{xy}{1+x+y}}{\ln 2} \leq \frac{xy}{\ln 2}. \tag{6.8}$$

Based on [36, Theorem 4.1], from $\mathbb{E}[R_n] < \infty$, we have $\sup_n \big\{\frac{\tau_d}{2} R_n - nc\big\} < \infty \ \ a.s.$; from $\mathbb{E}[R_n^2] < \infty$, we have $\mathbb{E}\left[\sup_n \big\{\frac{\tau_d}{2} R_n - nc\big\}\right] < \infty$. By decomposition similar to (43) in [33], the first condition for existence of an optimal stopping strategy can be proved. □

<u>Lemma</u> 6.2. The second condition is also satisfied, namely $\limsup_{n\to\infty} \{\frac{\tau_d}{2} R_n - \lambda(\tau_d + \sum_{l=1}^{n} t_l)\} = -\infty \ \ a.s.$

Proof. Using a similar method to that in [33], for $0 < \varepsilon < \tau_o$, we have the following decomposition:

$$\frac{\tau_d}{2} R_n - \lambda\big(\tau_d + \sum_{l=1}^{n} t_l\big) = \big[\frac{\tau_d}{2} R_n - n\lambda(\tau_o - \varepsilon) - \tau_d \lambda\big] + \big[\lambda \sum_{l=1}^{n} (\tau_o - \varepsilon - t_l)\big]. \tag{6.9}$$

From [36, Theorem 4.1], $\tau_o - \varepsilon > 0$, and (6.7), we have

$$\lim_{n\to\infty} \big[\frac{\tau_d}{2} R_n - n\lambda(\tau_o - \varepsilon)\big] = -\infty \ a.s. \tag{6.10}$$

Next we focus on the second component on the right-hand side of (6.9).

Using [36, Theorem 4.2], when $\mathbb{E}\big[\tau_o - \varepsilon - t_l\big] < 0$ holds, $\mathbb{E}\big[\sup_{n\geq 0} \sum_{l=1}^{n} (\tau_o - \varepsilon - t_l)\big] < \infty$ if and only if $\mathbb{E}\left[(\tau_o - \varepsilon - t_l)^+\right]^2 < \infty$, where $(\tau_o - \varepsilon - t_l)^+ = \max\{\tau_o - \varepsilon - t_l, 0\}$.

We have $\mathbb{E}[\tau_o - \varepsilon - t_l] = \tau_o - \varepsilon - \tau_o = -\varepsilon < 0$, and $\mathbb{E}[(\tau_o - \varepsilon - t_l)^+]^2 \leq \mathbb{E}[\tau_o - \varepsilon - t_l]^2 = (\tau_o - \varepsilon)^2 - 2(\tau_o - \varepsilon)\tau_o + \mathbb{E}[t_l^2] < \infty$ (since $\mathbb{E}[t_l^2]$ can be shown to be finite). As a result, we have $\mathbb{E}\big[\limsup_{n\to\infty} \big\{ \sum_{l=1}^{n} (\tau_o - \varepsilon - t_l)\big\}\big] \leq \mathbb{E}\big[\sup_{n\geq 0} \big\{ \sum_{l=1}^{n} (\tau_o - \varepsilon - t_l)\big\}\big] < \infty$
which leads to

$$\limsup_{n\to\infty} \big\{ \sum_{l=1}^{n} (\tau_o - \varepsilon - t_l)\big\} < \infty \ a.s. \tag{6.11}$$

From (6.9), (6.10), and (6.11), we have $\limsup_{n\to\infty} \big\{\frac{\tau_d}{2} R_n - \lambda(\tau_d + \sum_{l=1}^{n} t_l)\big\} = -\infty \ a.s.$
□

Based on Lemmas 6.1 and 6.2, the existence of an optimal stopping strategy is guaranteed.

<u>Theorem</u> 6.1. An optimal stopping strategy which achieves maximal system throughput $\sup_{N\geq 0}\frac{\mathbb{E}[Y_N]}{\mathbb{E}[T_N]}$ is given as: $N^* = \min\{n \geq 1 : R_n \geq 2\lambda^*\}$ where λ^* is the solution of the equation $\mathbb{E}\left[\max\left\{\frac{\tau_d}{2}R_n - \lambda\tau_d, 0\right\}\right] = \lambda\tau_o$.

Proof. Recall that to maximize throughput $\frac{\mathbb{E}[Y_N]}{\mathbb{E}[T_N]}$, we need to achieve $V^*(\lambda^*) = \sup_{N\geq 0}\{\mathbb{E}[Y_N] - \lambda^*\mathbb{E}[T_N]\}$ where λ^* satisfies $V^*(\lambda^*) = 0$. Here λ^* is actually maximal average throughput. Therefore we need to know expression of $V^*(\lambda)$.

For $\lambda \geq 0$, the stopping strategy which achieves maximal reward $V^*(\lambda)$ can be described as $N^* = \min\left\{n \geq 1 : \frac{\tau_d}{2}R_n - \lambda\tau_d \geq V^*(\lambda)\right\}$, where $V^*(\lambda)$ is determined by optimality equation

$$V_n^* = \max\left\{\frac{\tau_d}{2}R_n - \lambda\tau_d - \lambda\sum_{l=1}^{n} t_l,\ \mathbb{E}[V_{n+1}^*|X(1), ..., X(n)]\right\}. \tag{6.12}$$

Here V_n^* represents expected reward if the winner source at the nth observation does not stop and the optimal stopping strategy is followed starting from the $(n+1)$th observation. Since $V_n^* = V^*(\lambda) - \lambda\sum_{l=1}^{n-1} t_l$, after taking expectation over both sides of (6.12) we have:

$$\mathbb{E}[V^*(\lambda) - \lambda\sum_{l=1}^{n-1} t_l] = \mathbb{E}\big[\max\{\frac{\tau_d}{2}R_n - \lambda\tau_d - \lambda\sum_{l=1}^{n} t_l, V^*(\lambda) - \lambda\sum_{l=1}^{n} t_l\}\big] \tag{6.13}$$

which leads to

$$V^*(\lambda) = \mathbb{E}\big[\max\{\frac{\tau_d}{2}R_n - \lambda\tau_d, V^*(\lambda)\} - \lambda t_n\big].$$

Setting $V^*(\lambda^*) = 0$, the maximal throughput λ^* satisfies

$$\mathbb{E}[\max\{\frac{\tau_d}{2}R_n - \lambda^*\tau_d, 0\}] = \lambda^*\mathbb{E}[t_n] = \lambda^*\tau_o. \tag{6.14}$$

And an optimal stopping strategy which maximizes throughput is of form

$$N^* = \min\{n \geq 1 : \frac{\tau_d}{2}R_n - \lambda^*\tau_d \geq V^*(\lambda^*)\} = \min\{n \geq 1 : R_n \geq 2\lambda^*\}. \tag{6.15}$$

□

With threshold $2\lambda^*$ as a fixed value, our derived strategy N^* has a pure-threshold structure and achieves the maximal system throughput $\lambda^* = \frac{\mathbb{E}[Y_{N^*}]}{\mathbb{E}[T_{N^*}]}$. And as the solution of the equation $\mathbb{E}[\max\{\frac{\tau_d}{2}R_n - \lambda\tau_d, 0\}] = \lambda\tau_o$, the maximal system throughput λ^* always uniquely exists. The proof is similar to that of Proposition 3.1 in [33], and thus, is omitted. The uniqueness of λ^* is consistent with its physical meaning as the optimal system throughput.

With $\{R_n\}_{n=1,\ldots,\infty}$ i.i.d. and pure-threshold structure of N^*, the stopping time denoted N determined by the optimal stopping strategy N^* follows a geometric distribution with $\mathbb{P}(N = n) = F_{R_n}(2\lambda^*)^{n-1}\big(1 - F_{R_n}(2\lambda^*)\big)$ where $F_{R_n}(\cdot)$ means cumulative distribution function (CDF) of R_n given in (6.3). Let R_{N^*} denote the achievable rate when the winner source stops. It has the CDF as $F_{R_{N^*}}(x) = \mathbb{I}[x \geq 2\lambda^*]\frac{F_{R_n}(x)-F_{R_n}(2\lambda^*)}{1-F_{R_n}(2\lambda^*)}$.

With the stopping time N determined by the strategy N^* geometrically distributed, the expectation of the stopping time $\mathbb{E}[N] = \frac{1}{1-F_{R_n}(2\lambda^*)}$ is finite. According to Wald Theorem [36] we have $\mathbb{E}[T_N] = \mathbb{E}[t_l]\mathbb{E}[N] + \tau_d = \frac{\tau_o}{1-F_{R_n}(2\lambda^*)} + \tau_d$.

In addition, the pure-threshold structure largely simplifies implementation. In details, after the nth successful channel contention, Source $s(n)$ wins the channel and calculates its achievable transmission rate R_n (which is via the best relay). If $R_n \geq 2\lambda^*$, Source $s(n)$ transmits to the best relay node and the best relay node helps forward to Destination $s(n)$; otherwise, Source $s(n)$ gives up the transmission opportunity and re-contends for channel access with the other $(K-1)$ sources again. In this way, the maximal average system throughput λ^* can be achieved.

Note that the value of λ^* can be calculated off-line. And the following iterative algorithm can be used to calculate λ^*:

$$\lambda_{k+1} = \lambda_k + \alpha \cdot \Big\{\mathbb{E}\Big[\max\big\{\frac{\tau_d}{2}R_n - \lambda_k\tau_d, 0\big\}\Big] - \lambda_k\tau_o\Big\} \tag{6.16}$$

where λ_0 is a non-negative initial value and α is step size such that $\epsilon \leq \alpha \leq \frac{2-\epsilon}{\tau_o+\tau_d}$ where $\epsilon > 0$ can be arbitrarily selected.

<u>Theorem</u> 6.2. The sequence $\{\lambda_k\}$ generated by the iterative algorithm converges to λ^*.

Proof. Proposition 1.2.3 in [115] says that

"Let $\{x_k\}$ be a sequence generated by a gradient method $x_{k+1} = x_k + \alpha_k d_k$, where d_k is gradient related. Assume that for some constant $C > 0$ we have $\big\| \bigtriangledown h(x) - \bigtriangledown h(y)\big\| \leq C\big\|x - y\big\|, \forall x, y \in \Re$,[5] and that for all k we have $d_k \neq 0$ and $\epsilon \leq \alpha_k \leq (2-\epsilon)\beta$, where

[5]Note that this condition is called the *Lipschitz continuity condition*.

$\beta = \frac{\left\|\bigtriangledown h(x_k)'d_k\right\|}{C\left\|d_k\right\|^2}$ and ϵ is a fixed positive scalar. Then every limit point of x_k is a stationary point of h." Here $\bigtriangledown h(\cdot)$ means gradient of function $h(\cdot)$.

We take $\bigtriangledown h(\lambda) = \lambda\tau_o - \mathbb{E}\Big[\max\big\{\frac{\tau_d}{2}R_n - \lambda\tau_d, 0\big\}\Big]$. Then there is one unique solution satisfying $\bigtriangledown h(\lambda) = 0$, which is λ^* in our optimal stopping problem.

For the Lipschitz continuity condition, we have

$$
\begin{aligned}
|\bigtriangledown h(x) - \bigtriangledown h(y)| &= \big|x\tau_o - \mathbb{E}[\max\{\frac{\tau_d}{2}R_n - x\tau_d, 0\}] - y\tau_o + \mathbb{E}[\max\{\frac{\tau_d}{2}R_n - y\tau_d, 0\}]\big| \\
&\leq \big|x\tau_o - y\tau_o\big| + \Big|\mathbb{E}\big[\max\{\frac{\tau_d}{2}R_n - x\tau_d, 0\}\big] - \mathbb{E}\big[\max\{\frac{\tau_d}{2}R_n - y\tau_d, 0\}\big]\Big| \\
&\leq \tau_o|x-y| + \big|\mathbb{E}[\max\{\frac{\tau_d}{2}R_n - x\tau_d, 0\} - \max\{\frac{\tau_d}{2}R_n - y\tau_d, 0\}]\big| \\
&\leq (\tau_o + \tau_d)|x-y| = C|x-y|
\end{aligned}
$$

where $C = \tau_o + \tau_d$. This means the Lipschitz continuity condition in Proposition 1.2.3 in [115] is satisfied.

Define directions d_k as the steepest descent direction $d_k \triangleq -\bigtriangledown h(\lambda_k) = \mathbb{E}\big[\max\{\frac{\tau_d}{2}R_n - \lambda_k\tau_d, 0\}\big] - \lambda_k\tau_o$. Then it can be proved that $\{d_k\}$ is gradient related.

Then based on Proposition 1.2.3 in [115], a generated sequence $\{\lambda_k\}$ by

$$\lambda_{k+1} = \lambda_k + \alpha_k d_k \ \text{ when } \ \epsilon \leq \alpha \leq (2-\epsilon)\cdot\frac{1}{C} = \frac{2-\epsilon}{\tau_o + \tau_d} \tag{6.17}$$

converges to the stationary point of h, which is λ^*.

Iteration form (6.17) is actually the iteration form (6.16). □

6.3 Case II: with Partial CSI at a Winner Source

6.3.1 System Model

In the previous section, the winner source in each observation has CSI of links from itself to all relays and from all relays to its destination. Next we consider a more practical case that the winner source in each observation has only CSI of links from itself to relays. Since the winner source does not have CSI in the second hop, relay is not selected by the winner source. Rather, there is another channel access contention among the relays, with details as follows.

The channel contention of sources is similar to that in Section 6.2. The difference is as follows: If there is only one source, say Source i, transmitting RTS in a contention, there is no information exchange between relays and Destination i. So Source i has only its CSI to relays (obtained from the CTSs from the relays). And if Source i decides to stop, it broadcasts its packet to all relays, and then all relays start to contend for channel access, as follows. At the beginning of a time slot, each relay independently transmits an RTS with probability p_1. If no relay transmits RTS, or two or more relays transmit, then a new contention of relays is started subsequently. If only one relay, say Relay j, transmits RTS (in which information of Destination i is included), then Destination i estimates its channel gain with Relay j and replies with a CTS with channel gain information g_{ji} included. Then Relay j can decide 1) to stop (i.e., to forward its received packet to Destination i, and then a new source contention is started), or 2) to give up its transmission opportunity and then a new contention of relays is started.

The channel access is actually a bi-layer stopping problem: the main layer for channel access of sources, and the sub-layer for channel access of relays. In either layer, still define an observation as the process until a successful winner appears. So in the main layer, the winner source in the nth observation, denote $s(n)$, decides whether to stop based on its observed information $\left\{s(n), f_{s(n)1}(n), ..., f_{s(n)L}(n)\right\}$. In the sub-layer, the winner relay in the mth observation, denote $s(m)$, decides whether to stop based on its observed information $\{s(m), g_{s(m)s(n)}(m)\}$ and channel gain realization $f_{s(n)s(m)}(n)$ in the preceding first-hop transmission. Recall that information of $f_{s(n)j}(n)$ $(j = 1, 2, ..., L)$ is already obtained by Relay j when Source $s(n)$ broadcasts to relays in the first hop.

Similar to Section 6.2, the mean of duration of an observation in the main layer and the sub-layer are

$$\tau_o^s = \tau_{RTS} + L\tau_{CTS} + \frac{(1-p_0)^K}{Kp_0(1-p_0)^{K-1}} \cdot \delta + \frac{1-(1-p_0)^K - Kp_0(1-p_0)^{K-1}}{Kp_0(1-p_0)^{K-1}} \cdot \tau_{RTS}$$

$$\tau_o^r = \tau_{RTS} + \tau_{CTS} + \frac{(1-p_1)^L}{Lp_1(1-p_1)^{L-1}} \cdot \delta + \frac{1-(1-p_1)^L - Lp_1(1-p_1)^{L-1}}{Lp_1(1-p_1)^{L-1}} \cdot \tau_{RTS}.$$

Note that in this chapter, superscript ‘s’ and ‘r’ stand for source (first hop) and relays (second hop), respectively.

A winner source does not have CSI of links in the second hop (from relays to destinations). Rather, statistical information (e.g., channel gain distribution) of channel gains in the second hop is assumed to be available. Therefore, in the main layer, the reward (which is the source-to-destination data volume) in the nth observation is the expected reward in the sub-layer. On the other hand, in the sub-layer, the stopping problem should be conditioned on channel gain realization of the preceding first-hop transmission.

In the main layer, let n and N denote the observation index and stopping time, respectively. And in the sub-layer, let m and M denote the observation index and stopping time, respectively. We use $\mathbb{E}_1[\cdot]$ and $\mathbb{E}_2[\cdot]$ to present expectations on the main layer and sub-layer, respectively.

6.3.2 Intuitive Stopping strategy

An intuitive method to solve the bi-layer stopping problem is to let the sub-layer and main layer apply optimal stopping theory to maximize sub-layer and main-layer throughput, respectively.

We first consider the sub-layer. The relays already know channel gain realization in the preceding first-hop transmission of $\mathcal{F} = \big\{f_{s(n)1}(n), ..., f_{s(n)L}(n)\big\}$.[6] Then in the mth observation, the achievable rate of the winner relay, $s(m)$, is

$$R_m = \sum_{j=1}^{L} \mathbb{I}[s(m) = j] \log_2 \big(1 + \frac{P_s P_r |f_{s(n)j}(n)|^2 |g_{js(n)}(m)|^2}{1 + P_s |f_{s(n)j}(n)|^2 + P_r |g_{js(n)}(m)|^2}\big). \tag{6.18}$$

The reward in the mth observation is $Y_m = \frac{\tau_d}{2} R_m$. The cost is the total waiting time until the mth observation plus the data transmission time in the second hop: $T_m = \sum_{l=1}^{m} t_l^r + \frac{\tau_d}{2}$, where t_l^r is the time used in the lth observation. Then we need to find an optimal stopping rule M^* in the sub-layer to attain the maximal $\lambda^* = \sup\limits_{M \geq 0} \frac{\mathbb{E}_2[Y_M|\mathcal{F}]}{\mathbb{E}_2[T_M|\mathcal{F}]}$.

In the main layer, define T_n as the total waiting time until the nth observation plus the data transmission time in the first hop: $T_n = \sum_{l=1}^{n} t_l^s + \frac{\tau_d}{2}$, where t_l^s is the time used in the lth observation. If the stopping time is N, then the reward is $\mathbb{E}_2[Y_{M^*}|\mathcal{F}]$, and the waiting time is $\mathbb{E}_2[T_{M^*}|\mathcal{F}] + T_N$. Then we need to find an optimal stopping rule N^* to attain the maximal $\sup\limits_{N \geq 0} \frac{\mathbb{E}_1[\mathbb{E}_2[Y_{M^*}|\mathcal{F}]]}{\mathbb{E}_1[\mathbb{E}_2[T_{M^*}|\mathcal{F}] + T_N]}$.

For the sub-layer optimal stopping problem, we have the following theorem.

[6] Note that it means Relay j knows $f_{s(n)j}(n)$, $j = 1, 2, ..., L$.

Theorem 6.3. Conditioned on $\mathcal{F}$, a sub-layer optimal stopping rule achieving the maximal sub-layer throughput $\lambda^* = \sup\limits_{M \geq 0} \frac{\mathbb{E}_2[Y_M|\mathcal{F}]}{\mathbb{E}_2[T_M|\mathcal{F}]}$ is given as: $M^* = \min\{m \geq 1 : R_m \geq \lambda^*\}$ where λ^* is the unique solution of the equation $\mathbb{E}_2[\max\{R_m - \lambda, 0\}|\mathcal{F}] = \frac{2\lambda\tau_o^r}{\tau_d}$ and always exists.

Proof. We first prove the finiteness of $\mathbb{E}[R_m^2]$.

$$\begin{aligned}
\mathbb{E}_2[R_m^2|\mathcal{F}] &= \mathbb{E}_2\big[\sum_{j=1}^{L} \mathbb{I}[s(m)=j] \log_2^2\big(1+\tfrac{P_sP_r|f_{s(n)j}(n)|^2|g_{js(n)}(m)|^2}{1+P_s|f_{s(n)j}(n)|^2+P_r|g_{js(n)}(m)|^2}\big)\big|\mathcal{F}\big] \\
&= \sum_{j=1}^{L} \tfrac{1}{L}\mathbb{E}_2\big[\log_2^2\big(1+\tfrac{P_sP_r|f_{s(n)j}(n)|^2|g_{js(n)}(m)|^2}{1+P_s|f_{s(n)j}(n)|^2+P_r|g_{js(n)}(m)|^2}\big)\big|\mathcal{F}\big] \\
&\overset{(e)}{\leq} \sum_{j=1}^{L} \tfrac{1}{L}\tfrac{1}{(\ln 2)^2} P_r^2 \mathbb{E}[|g_{js(n)}|^4] \\
&= \sum_{j=1}^{L} \tfrac{1}{L}\tfrac{2}{(\ln 2)^2} P_r^2 \sigma_g^4 < \infty
\end{aligned} \tag{6.19}$$

where (e) comes from the fact that for $x, y \geq 0$, we have

$$\log_2\left(1+\frac{xy}{1+x+y}\right) \leq \frac{\frac{xy}{1+x+y}}{\ln 2} \leq \frac{y}{\ln 2}. \tag{6.20}$$

With the finite property of $\mathbb{E}_2[R_m^2|\mathcal{F}]$, we have

$$\mathbb{E}_2[R_m|\mathcal{F}] < \infty. \tag{6.21}$$

Then similar to proofs of Lemmas 6.1 and 6.2, the existence conditions of an optimal stopping rule in the sub-layer can be proved. With the reward as $\frac{\tau_d}{2}R_m - \lambda\frac{\tau_d}{2} - \lambda\sum_{l=1}^{m} t_l^r$, by following a similar way to that in proof of Theorem 6.1, we can obtain an optimal stopping rule for the sub-layer as the form: $M^* = \min\{m \geq 1 : R_m \geq \lambda^*\}$ where λ^* satisfies the equality

$$\mathbb{E}_2[\max\{R_m - \lambda, 0\}|\mathcal{F}] = \frac{2\lambda\tau_o^r}{\tau_d}. \tag{6.22}$$

And the existence and uniqueness of λ^* can be straightforwardly proved. □

Define $F_{R_m}(\cdot)$ as the CDF of R_m given in (6.18). The sub-layer optimal stopping rule has the following property.

Corollary 6.1. Conditioned on $\mathcal{F}$, we have finite λ^*, $\mathbb{E}_2[T_{M^*}|\mathcal{F}] = \frac{\tau_o^r}{1-F_{R_m}(\lambda^*)} + \frac{\tau_d}{2}$ and $\mathbb{E}_2[Y_{M^*}|\mathcal{F}] = \frac{\lambda^*\tau_o^r}{1-F_{R_m}(\lambda^*)} + \frac{\lambda^*\tau_d}{2}$.

Proof. $\mathbb{E}_2[\max\{R_m - \lambda, 0\}|\mathcal{F}]$ is a decreasing function from $+\infty$ to 0 with respect to λ, and $\frac{\lambda\tau_o^r}{\tau_d}$ linearly increases with respect to λ. Hence, the uniqueness and non-negativeness of the root λ^* are guaranteed, since λ^* is the root of $\mathbb{E}_2[\max\{R_m - \lambda, 0\}|\mathcal{F}] = \frac{2\lambda\tau_o^r}{\tau_d}$.

Further, we have

$$\begin{aligned} \frac{2\lambda^*\tau_o^r}{\tau_d} &= \mathbb{E}_2[\max\{R_m - \lambda^*, 0\}|\mathcal{F}] \\ &\leq \mathbb{E}_2[R_m|\mathcal{F}] \\ &\overset{\text{from (6.21)}}{<} \infty \end{aligned} \tag{6.23}$$

which leads to $\lambda^* < \infty$.

Stopping time M in the sub-layer is geometrically distributed. Then according to Wald Theorem [36], $\mathbb{E}_2[T_{M^*}|\mathcal{F}] = \frac{\tau_o^r}{1-F_{R_m}(\lambda^*)} + \frac{\tau_d}{2}$. Also, we have $\mathbb{E}_2[Y_{M^*}|\mathcal{F}] = \frac{\lambda^*\tau_o^r}{1-F_{R_m}(\lambda^*)} + \frac{\lambda^*\tau_d}{2}$. □

Based on the acquired strategy M^* for the sub-layer stopping problem, a main-layer optimal stopping rule which achieves maximal system throughput is given in the following theorem.

Theorem 6.4. An optimal stopping rule for the main-layer problem is of the form $N^* = \min\{n \geq 1 : R_n^1 - \gamma^* R_n^2 \geq \gamma^* \frac{\tau_d}{2}\}$ where γ^* satisfies the equation $\mathbb{E}_1\big[\max\{R_n^1 - \gamma R_n^2 - \gamma\frac{\tau_d}{2}, 0\}\big] = \gamma\tau_o^s$, and R_n^1 and R_n^2 are given as: $R_n^1 = \lambda^*\mathbb{E}_2[T_{M^*}|\mathcal{F}]$ and $R_n^2 = \mathbb{E}_2[T_{M^*}|\mathcal{F}]$.[7]

Proof. Recall that to maximize throughput $\frac{\mathbb{E}_1[\lambda^*\mathbb{E}_2[T_{M^*}|\mathcal{F}]]}{\mathbb{E}_1[\mathbb{E}_2[T_{M^*}|\mathcal{F}]+T_{N^*}]}$, we need to achieve

$$V^*(\gamma^*) = \sup_{N\geq 0}\big\{\mathbb{E}_1\big[\lambda^*\mathbb{E}_2[T_{M^*}|\mathcal{F}] - \gamma^*\big(\mathbb{E}_2[T_{M^*}|\mathcal{F}] + \frac{\tau_d}{2} + \sum_{l=1}^{N} t_l^s\big)\big]\big\} \tag{6.24}$$

where γ^* satisfies $V^*(\gamma^*) = 0$. To derive an optimal stopping rule, we first need to calculate $V^*(\gamma)$.

For $\gamma \geq 0$, an optimal stopping rule

$$N^*(\gamma) = \min\{n \geq 1 : R_n^1 - \gamma\frac{\tau_d}{2} - \gamma R_n^2 \geq V^*(\gamma)\} \tag{6.25}$$

[7]Note that here M^* is the optimal stopping rule of the sub-layer conditioned on $\mathcal{F}$, and λ^* is the corresponding maximal throughput in the sub-layer stopping problem. Therefore, R_n^1 and R_n^2 are functions of $\mathcal{F}$.

exists (with proof given in Appendix 6.A.) and achieves $V^*(\gamma)$ which satisfies the equation:

$$\mathbb{E}_1\big[V^*(\gamma)-\gamma\sum_{l=1}^{n-1}t_l^s\big] = \mathbb{E}_1\Big[\max\big\{R_n^1-\gamma\frac{\tau_d}{2}-\gamma R_n^2-\gamma\sum_{l=1}^{n}t_l^s,\ V^*(\gamma)-\gamma\sum_{l=1}^{n}t_l^s\big\}\Big] \quad (6.26)$$

which leads to

$$V^*(\gamma) = \mathbb{E}_1[\max\{R_n^1 - \gamma\frac{\tau_d}{2} - \gamma R_n^2,\ V^*(\gamma)\} - \gamma t_n^s]. \quad (6.27)$$

Setting $V^*(\gamma^*) = 0$ in (6.27), the maximal throughput γ^* satisfies

$$\mathbb{E}_1\big[\max\{R_n^1 - \gamma^* R_n^2 - \gamma^*\frac{\tau_d}{2}, 0\}\big] = \tau_o^s. \quad (6.28)$$

And an optimal stopping rule which achieves γ^* is $N^* = \min\{n \geq 1 : R_n^1 - \gamma^* R_n^2 \geq \gamma^*\frac{\tau_d}{2}\}$. □

Note that here γ^* is actually the maximal main-layer system throughput.

From Theorems 6.3 and 6.4, it can be seen that, the intuitive optimal stopping strategy $\{N^*, M^*\}$ with $M^* = \min\{m \geq 1 : R_m \geq \lambda^*\}$ and $N^* = \min\{n \geq 1 : R_n^1 - \gamma^* R_n^2 \geq \gamma^*\frac{\tau_d}{2}\}$ has semi-pure-threshold structure. In details, with sub-layer stopping rule M^*, its threshold is not a fixed value, but depends on channel gain realization $\mathcal{F}$ in the preceding first-hop transmission. Different from M^*, the main-layer stopping rule N^* has a fixed-valued threshold $\gamma^*\frac{\tau_d}{2}$.

The intuitive stopping strategy can be implemented as follows.

For channel access of sources, upon a successful contention in the nth observation, the winner source, $s(n)$, has the information of its channel gains $\mathcal{F}=\{f_{s(n)1}(n), ..., f_{s(n)L}(n)\}$. Source $s(n)$ can calculate R_n^1 and R_n^2 by solving the sub-layer optimal stopping problem conditioned on $\mathcal{F}$. During the calculation of R_n^1 and R_n^2, Source $s(n)$ needs to calculate λ^*, which is the threshold of the sub-layer optimal stopping rule conditioned on $\mathcal{F}$. In the main-layer stopping rule, γ^* is a fixed value satisfying $\mathbb{E}_1\big[\max\{R_n^1 - \gamma R_n^2 - \gamma\frac{\tau_d}{2}, 0\}\big] = \gamma\tau_o^s$.

- If $R_n^1 - \gamma^* R_n^2 < \gamma^*\frac{\tau_d}{2}$, Source $s(n)$ gives up its transmission opportunity and re-contend with other sources.

- If $R_n^1 - \gamma^* R_n^2 \geq \gamma^*\frac{\tau_d}{2}$, Source $s(n)$ broadcasts its data and the value of λ^* to all relays, and the channel contention of relays starts. Upon a successful contention in the mth observation, the winner relay, $s(m)$, which has information of $f_{s(n)s(m)}(n)$ in the preceding first-hop transmission, calculates its source-to-destination rate R_m. If $R_m < \lambda^*$, Relay $s(m)$ gives up its transmission opportunity, and re-contends with other relays. Otherwise, Relay $s(m)$ forwards its received data (from Source $s(n)$) to Destination $s(n)$, and the source-to-destination transmission process for the packet from Source $s(n)$ is complete, and all source nodes start a new contention.

Note that, the threshold in the main layer γ^* (for simplicity of presentation, the constant factor $\frac{\tau_d}{2}$ is omitted) can be calculated off-line, while the threshold λ^* in the sub-layer depends on the channel gain realization $\mathcal{F}$ in the preceding first-hop transmission, and thus, should be calculated online at Source $s(n)$, who knows $\mathcal{F}$. The following iterative algorithm can be used to calculate γ^* and λ^*.

To calculate λ^*, we have

$$\lambda_{l+1} = \lambda_l + \alpha_\lambda \cdot \left\{\mathbb{E}_2[\max\{R_m - \lambda_l, 0\}|\mathcal{F}] - \frac{2\lambda_l \tau_o^r}{\tau_d}\right\} \tag{6.29}$$

where step size α_λ satisfies $\epsilon \leq \alpha_\lambda \leq \frac{\tau_d(2-\epsilon)}{2\tau_o^r + \tau_d}$ for a fixed positive ϵ.

For main-layer problem, to calculate γ^*, we have

$$\gamma_{k+1} = \gamma_k + \alpha_\gamma \cdot \left\{\mathbb{E}_1\left[\max\{R_n^1 - \gamma_k R_n^2 - \gamma_k \frac{\tau_d}{2}, 0\}\right] - \gamma_k \tau_o^s\right\} \tag{6.30}$$

where step size α_γ satisfies $\epsilon \leq \alpha_\gamma \leq \frac{2(2-\epsilon)}{2\mathbb{E}_1[R_n^2] + \tau_d + 2\tau_o^s}$ for a fixed positive ϵ.

Theorem 6.5. The sequence $\{\gamma_k\}$ generated by the iterative algorithm converges to γ^*.

Proof. Similar to proof of Theorem 6.2, Lipschitz continuity conditions in the sub-layer and main layer are derived as follows.

In the sub-layer problem, we have:

$$\begin{aligned}
&\left|\frac{2x\tau_o^r}{\tau_d} - \mathbb{E}_2[\max\{R_m - x, 0\}|\mathcal{F}] - \frac{2y\tau_o^r}{\tau_d} + \mathbb{E}_2[\max\{R_m - y, 0\}|\mathcal{F}]\right| \\
&\quad \leq \frac{2\tau_o^r}{\tau_d}|x-y| + \left|\mathbb{E}_2[\max\{R_m - x, 0\} - \max\{R_m - y, 0\}|\mathcal{F}]\right| \\
&\quad \leq (\frac{2\tau_o^r}{\tau_d} + 1)|x-y|.
\end{aligned}$$

Step-size α_λ is fixed, which satisfies $\epsilon \leq \alpha_\lambda \leq \frac{\tau_d(2-\epsilon)}{2\tau_o^r + \tau_d}$.

In the main-layer problem, we have:

$$\begin{aligned}
&\left|x\tau_o^s - \mathbb{E}_1[\max\{R_n^1 - xR_n^2 - x\frac{\tau_d}{2}, 0\}] - y\tau_o^s + \mathbb{E}_1[\max\{R_n^1 - yR_n^2 - y\frac{\tau_d}{2}, 0\}]\right| \\
&\quad \leq \tau_o^s|x-y| + \left|\mathbb{E}_1[\max\{R_n^1 - xR_n^2 - x\frac{\tau_d}{2}, 0\} - \max\{R_n^1 - yR_n^2 - y\frac{\tau_d}{2}, 0\}]\right| \\
&\quad \leq (\tau_o^s + \mathbb{E}_1[R_n^2] + \frac{\tau_d}{2}) \cdot |x-y|.
\end{aligned}$$

Step-size α_γ is fixed, which satisfies $\epsilon \leq \alpha_\gamma \leq \frac{2(2-\epsilon)}{2\mathbb{E}_1[R_n^2]+\tau_d+2\tau_o^s}$. □

Since the calculation of γ^* involves the calculation of λ^* conditioned on $\mathcal{F}$, convergence of $\{\gamma_k\}$ to γ^* also guarantees convergence of $\{\lambda_l\}$ to λ^*.

6.3.3 Non-optimality of Intuitive Stopping strategy

The intuitive stopping strategy $\{N^*, M^*\}$ first maximizes sub-layer system throughput and then maximizes that of main-layer system. It is interesting to notice that the intuitive stopping strategy is not optimal, as follows.

The expected system throughput can be expressed as $\frac{\mathbb{E}_1[\lambda^*\mathbb{E}_2(T_{M^*}|\mathcal{F})]}{\mathbb{E}_1[\mathbb{E}_2[T_{M^*}|\mathcal{F}]+T_{N^*}]}$ in the intuitive stopping strategy. The sub-layer stopping rule M^* maximizes λ^*. Considering the term T_{N^*} in the expression of the expected system throughput, the sub-layer stopping rule M^*, which maximizes λ^*, may not maximize $\frac{\mathbb{E}_1[\lambda^*\mathbb{E}_2[T_{M^*}|\mathcal{F}]]}{\mathbb{E}_1[\mathbb{E}_2[T_{M^*}|\mathcal{F}]+T_{N^*}]}$.

6.3.4 Optimal Stopping strategy

Next we derive an optimal stopping strategy for the sub-layer and main layer.

For $\gamma \geq 0$ and a particular stopping rule in the sub-layer (which is conditioned on $\mathcal{F}$) denoted M, the maximal average reward achieved by main-layer optimal stopping rule can be expressed as:

$$V^*(\gamma) := \sup_{N\geq 0} \left\{\mathbb{E}_1\left[\mathbb{E}_2[Y_M|\mathcal{F}] - \gamma(\mathbb{E}_2[T_M|\mathcal{F}] + T_N)\right]\right\} \tag{6.31}$$

which is equivalent to

$$V^*(\gamma) := \sup_{N\geq 0} \left\{\mathbb{E}_1\left[\mathbb{E}_2[Y_M - \gamma T_M|\mathcal{F}] - \gamma T_N\right]\right\}. \tag{6.32}$$

In the expression of (6.32), the sub-layer affects only the term $\mathbb{E}_2[Y_M - \gamma T_M|\mathcal{F}]$. Therefore, to increase the maximal system throughput γ^*, we need to increase $V^*(\gamma)$ (this is because $V^*(\gamma)$ is a decreasing function of γ, and γ^* is the root of $V^*(\gamma) = 0$). And to achieve the largest $V^*(\gamma)$, the sub-layer should maximize $\mathbb{E}_2[Y_M - \gamma T_M|\mathcal{F}]$. Based on this, we have the following theorem for the sub-layer. Here we use $W^*(\gamma)$ to denote the maximal reward $\sup_{M\geq 0} \mathbb{E}_2[Y_M - \gamma T_M|\mathcal{F}]$ in the sub-layer.

<u>Theorem</u> 6.6. For fixed $\gamma \geq 0$, an optimal stopping rule $M^*(\gamma)$ for maximizing $\mathbb{E}_2[Y_M - \gamma T_M|\mathcal{F}]$ is of the form: $M^*(\gamma) = \min\{m \geq 1 : \frac{\tau_d}{2} R_m \geq W^*(\gamma) + \frac{\tau_d}{2}\gamma\}$ where $W^*(\gamma)$ satisfies

$$\mathbb{E}_2\big[\max\{\frac{\tau_d}{2} R_m - \frac{\tau_d}{2}\gamma, W^*(\gamma)\}|\mathcal{F}\big] = W^*(\gamma) + \gamma\tau_o^r. \tag{6.33}$$

Proof. Similar to proof of (6.19), we have $\mathbb{E}_2[(R_m)^2] < \infty$, which guarantees existence of an optimal stopping rule. To achieve maximal reward $W^*(\gamma) = \sup_{M\geq 0}\{\mathbb{E}_2[Y_M - \gamma T_M|\mathcal{F}]\}$, an optimal stopping rule takes the form: $M^*(\gamma) = \min\{m \geq 1 : \frac{\tau_d}{2} R_m \geq W^*(\gamma) + \frac{\tau_d}{2}\gamma\}$ where $W^*(\gamma)$ satisfies the equation

$$\mathbb{E}_2\big[\max\{\frac{\tau_d}{2} R_m - \frac{\tau_d}{2}\gamma, W^*(\gamma)\}|\mathcal{F}\big] = W^*(\gamma) + \gamma\tau_o^r. \tag{6.34}$$

Rearranging terms in (6.34), we have

$$\mathbb{E}_2\big[\max\{\frac{\tau_d}{2} R_m - \frac{\tau_d}{2}\gamma - W^*(\gamma), 0\}\big|\mathcal{F}\big] = \gamma\tau_o^r. \tag{6.35}$$

Since the left hand side of (6.35) continuously decreases from ∞ to 0 with $W^*(\gamma)$, while the right hand side is a constant, a finite unique solution $W^*(\gamma)$ always exists. □

Although Theorem 6.6 is for any particular value of γ, it is desired that the sub-layer stopping rule is corresponding to the maximal system throughput γ^*. How to obtain the value of γ^* will be discussed in the main-layer stopping rule, as follows.

<u>Theorem</u> 6.7. With the sub-layer system following the strategy $M^*(\gamma^*)$, an optimal strategy to maximize the average system throughput is given as $N^* = \min\{n \geq 1 : W^*(\gamma^*) \geq \frac{\tau_d}{2}\gamma^*\}$ where γ^* satisfies $\mathbb{E}_1[\max\{W^*(\gamma) - \frac{\tau_d}{2}\gamma, 0\}] = \gamma\tau_o^s$.

Proof. Recall that to maximize throughput $\frac{\mathbb{E}_1[\mathbb{E}_2[Y_{M^*}|\mathcal{F}]]}{\mathbb{E}_1[\mathbb{E}_2[T_{M^*}|\mathcal{F}]+T_{N^*}]}$, we need to achieve

$$V^*(\gamma^*) = \sup_{N\geq 0}\{\mathbb{E}_1\big[W^*(\gamma^*) - \gamma^*(\frac{\tau_d}{2} + \sum_{l=1}^{N} t_l^s)\big]\} \tag{6.36}$$

where γ^* satisfies $V^*(\gamma^*) = 0$.

To derive an optimal stopping rule, we first need to calculate $V^*(\gamma)$.

For $\gamma \geq 0$, an optimal stopping rule N^* to achieve $V^*(\gamma)$ exists which is proved as follows.

Similar to (6.19), conditioned on $\mathcal{F}$, we have

$$\begin{aligned} R_m &= \sum_{j=1}^{L} \mathbb{I}[s(m) = j] \log_2 \left(1 + \tfrac{P_s P_r |f_{s(n)j}(n)|^2 |g_{js(n)}(m)|^2}{1+P_s|f_{s(n)j}(n)|^2+P_r|g_{js(n)}(m)|^2}\right) \\ &\leq \max_{j\in\{1,\ldots,L\}} \left\{ \log_2 \left(1 + \tfrac{P_s P_r |f_{s(n)j}(n)|^2 |g_{js(n)}(m)|^2}{1+P_s|f_{s(n)j}(n)|^2+P_r|g_{js(n)}(m)|^2}\right)\right\} \\ &\leq \tfrac{P_s}{\ln 2} \max_{j\in\{1,\ldots,L\}} |f_{s(n)j}(n)|^2. \end{aligned} \tag{6.37}$$

From (6.35), we have $W^*(\gamma) < \frac{\tau_d}{2} \cdot \frac{P_s}{\ln 2} \max_{j\in\{1,\ldots,L\}} |f_{s(n)j}(n)|^2$, which leads to $\mathbb{E}_1[(W^*(\gamma))^2] < \infty$ by integrating $(W^*(\gamma))^2$ over joint PDF of $\{|f_{s(n)1}(n)|^2, ..., |f_{s(n)L}(n)|^2\}$ where $f_{s(n)1}(n), f_{s(n)2}(n), ..., f_{s(n)L}(n)$ are exponentially distributed i.i.d. random variables.

Similar to proofs of Lemmas 6.1 and 6.2, $\mathbb{E}_1[W^*(\gamma)^2] < \infty$ and $\mathbb{E}_1[(t_l^s)^2] < \infty$ guarantee existence of an optimal stopping rule.

By using optimal stopping rule $N^*(\gamma) = \{n \geq 1 : W^*(\gamma) - \frac{\tau_d}{2}\gamma \geq V^*(\gamma)\}$, we can achieve $V^*(\gamma)$ which satisfies the equation as

$$\mathbb{E}_1[\max\{W^*(\gamma) - \frac{\tau_d}{2}\gamma, V^*(\gamma)\}] = V^*(\gamma) + \gamma\tau_o^s.$$

Setting $V^*(\gamma) = 0$, the maximal throughput γ^* satisfies

$$\mathbb{E}_1[\max\{W^*(\gamma^*) - \frac{\tau_d}{2}\gamma^*, 0\}] = \gamma^*\tau_o^s.$$

And an optimal stopping rule which maximizes the throughput is $N^* = \min\{n \geq 1 : W^*(\gamma^*) \geq \frac{\tau_d}{2}\gamma^*\}$. □

Overall, we can see that the optimal stopping strategy $\{N^*, M^*\}$ has the form of $M(\gamma^*) = \min\{m \geq 1 : \frac{\tau_d}{2} R_m \geq W^*(\gamma^*) + \frac{\tau_d}{2}\gamma^*\}$ and $N^* = \{n \geq 1 : W^*(\gamma^*) \geq \frac{\tau_d}{2}\gamma^*\}$, which achieves average system throughput maximum γ^*. Here γ^* is a fixed value satisfying

$$\mathbb{E}_1[\max\{W^*(\gamma) - \frac{\tau_d}{2}\gamma, 0\}] = \gamma\tau_o^s \tag{6.38}$$

where $W^*(\gamma)$ is an unique root of $\mathbb{E}_2[\max\{\frac{\tau_d}{2}R_m - \frac{\tau_d}{2}\gamma, W^*(\gamma)\}|\mathcal{F}] = W^*(\gamma) + \gamma\tau_o^r$.

Note that the optimal stopping strategy $\{N^*, M^*\}$ has also semi-pure-threshold structure, as in the main layer the threshold $\frac{\tau_d}{2}\gamma^*$ is a fixed value, while in the sub-layer the threshold $W^*(\gamma^*) + \frac{\tau_d}{2}\gamma^*$ is conditioned on the channel gain realization in the preceding first-hop transmission.

The optimal stopping strategy can be carried out as follows.

For channel access of sources, upon a successful contention in the nth observation, the winner source, $s(n)$, has the information of its channel gains $\mathcal{F} = \{f_{s(n)1}(n), ..., f_{s(n)L}(n)\}$. Source $s(n)$ can calculate $W^*(\gamma^*)$ by solving the sub-layer optimal stopping problem conditioned on $\mathcal{F}$.

- If $W^*(\gamma^*) < \frac{\tau_d}{2}\gamma^*$, Source $s(n)$ gives up its transmission opportunity and re-contend with other sources.

- If $W^*(\gamma^*) \geq \frac{\tau_d}{2}\gamma^*$, Source $s(n)$ broadcasts its data and also the value of $W^*(\gamma^*) + \frac{\tau_d}{2}\gamma^*$ to all relays, and channel contention of relays starts. Upon a successful contention in the mth observation, the winner relay, $s(m)$, who has information of $f_{s(n)s(m)}(n)$ in the preceding first-hop transmission, calculates its source-to-destination rate R_m. If $\frac{\tau_d}{2}R_m < W^*(\gamma^*) + \frac{\tau_d}{2}\gamma^*$, Relay $s(m)$ gives up its transmission opportunity, and re-contends with other relays; otherwise, Relay $s(m)$ forwards its received data (from Source $s(n)$ in the preceding first-hop transmission) to Destination $s(n)$, and the source-to-destination transmission process for the packet from Source $s(n)$ is complete, and all source nodes start a new contention.

Similar to the intuitive stopping strategy, the threshold in the main layer γ^* (with the constant factor $\frac{\tau_d}{2}$ omitted) can be calculated off-line, while the threshold $W^*(\gamma^*)$ (with the constant $\frac{\tau_d}{2}\gamma^*$ omitted) is dependent on $\mathcal{F}$, and thus, should be calculated online at Source $s(n)$, who knows $\mathcal{F}$. The following iterative algorithm can be used to calculate γ^* and $W^*(\gamma^*)$.

In the main layer, iterative algorithm is given below:

$$\gamma_{k+1} = \gamma_k + \alpha_\gamma(\mathbb{E}_1\big[\max\{W^*(\gamma_k) - \frac{\tau_d}{2}\gamma_k, 0\}\big] - \gamma_k\tau_o^s) \tag{6.39}$$

where step size α_γ satisfies $\epsilon \leq \alpha_\gamma \leq \frac{2-\epsilon}{\tau_d + \tau_o^s + \tau_o^r \cdot \mathbb{E}_1[\frac{1}{1-F_{R_m}(2W^*(0)/\tau_d)}]}$ for a fixed positive ϵ.

For each iteration of main layer, $W^*(\gamma_k)$ can be calculate below:

$$W_{l+1}(\gamma_k) = W_l(\gamma_k) + \alpha_W\Big(\mathbb{E}_2\big[\max\{\frac{\tau_d}{2}R_m - \frac{\tau_d}{2}\gamma_k - W_l(\gamma_k), 0\}\big|\mathcal{F}\big] - \gamma_k\tau_o^r\Big) \quad (6.40)$$

where step size α_W satisfies $\epsilon \le \alpha_W \le 2-\epsilon$ for a fixed positive ϵ.

Theorem 6.8. The sequence $\{\gamma_k\}$ generated by the iterative algorithm converges to γ^*.

Proof. Similar to proof of Theorem 6.2, Lipschitz continuity conditions in the sub-layer and main layer are derived as follows.

In the sub-layer problem, we have:

$$\begin{aligned}
&\big|\gamma\tau_o^r - \mathbb{E}_2[\max\{\frac{\tau_d}{2}R_m - \frac{\tau_d}{2}\gamma - x, 0\}\big|\mathcal{F}] - \gamma\tau_o^r + \mathbb{E}_2[\max\{\frac{\tau_d}{2}R_m - \frac{\tau_d}{2}\gamma - y, 0\}\big|\mathcal{F}]\big| \\
&\quad \le \big|\mathbb{E}_2[\max\{\frac{\tau_d}{2}R_m - \frac{\tau_d}{2}\gamma - x, 0\} - \max\{\frac{\tau_d}{2}R_m - \frac{\tau_d}{2}\gamma - y, 0\}\big|\mathcal{F}]\big| \\
&\quad \le |x-y|.
\end{aligned}$$

So step-size α_W is determined to satisfy $\epsilon \le \alpha_W \le 2-\epsilon$.

Conditioned on $\mathcal{F}$, we have property of $W^*(\gamma) = \sup_{M>0} \mathbb{E}_2[Y_M - \gamma T_M|\mathcal{F}]$ as follows. Without loss of generality, assume $x > y$. We have

$$W^*(y) - \mathbb{E}_2[T_{M(y)}](x-y) = \mathbb{E}_2[Y_{M(y)}] - x\mathbb{E}_2[T_{M(y)}] \le W^*(x)$$

which leads to

$$W^*(y) - W^*(x) \le \mathbb{E}_2[T_{M(y)}](x-y). \quad (6.41)$$

Consider main-layer problem, we have:

$$\begin{aligned}
&\big|x\tau_o^s - \mathbb{E}_1\big[\max\{W^*(x) - \frac{\tau_d}{2}x, 0\}\big] - y\tau_o^s + \mathbb{E}_1[\max\{W^*(y) - \frac{\tau_d}{2}y, 0\}]\big| \\
&\le \tau_o^s|x-y| + \big|\mathbb{E}_1[\max\{W^*(x) - \frac{\tau_d}{2}x, 0\} - \max\{W^*(y) - \frac{\tau_d}{2}y, 0\}]\big| \\
&\overset{\text{from (6.41)}}{\le} \tau_o^s|x-y| + \frac{\tau_d}{2}|x-y| + \mathbb{E}_1[\mathbb{E}_2(T_{M(y)}|\mathcal{F})]|x-y| \\
&\le (\tau_o^s + \frac{\tau_d}{2})|x-y| + \mathbb{E}_1\big[\frac{\tau_d}{2} + \tau_o^r \cdot \frac{1}{1 - F_{R_m}(y + \frac{2}{\tau_d}W^*(y))}\big]|x-y|. \quad (6.42)
\end{aligned}$$

From Equation (6.33), we have

$$y\tau_o^r = \mathbb{E}_2\big[\max\{\frac{\tau_d}{2}R_m - \frac{\tau_d}{2}y - W^*(y), 0\}\big] \tag{6.43}$$

which means that $\frac{\tau_d}{2}y + W^*(y)$ is a decreasing function of $y \in \Re^+$ with maximum at $y = 0$.

Similar to proof of Theorem 6.4, for $\forall y \in \Re^+$, we have $\mathbb{E}_1[\frac{1}{1-F_{R_m}(y+\frac{2}{\tau_d}W^*(y))}]$ finitely bounded. Also, using Monotone Convergence Theorem we have:

$$\lim_{y\to\infty} \mathbb{E}_1\big[\frac{1}{1 - F_{R_m}(y + \frac{2}{\tau_d}W^*(y))}\big] = \mathbb{E}_1\big[\lim_{y\to\infty} \frac{1}{1 - F_{R_m}(y + \frac{2}{\tau_d}W^*(y))}\big] = 1.$$

So if we define $C = \tau_d + \tau_o^s + \tau_o^r \cdot \mathbb{E}_1[\frac{1}{1-F_{R_m}(2W^*(0)/\tau_d)}]$, then from (6.42) we have

$$\big|x\tau_o^s - \mathbb{E}_1[\max\{W^*(x) - \frac{\tau_d}{2}x, 0\}] - y\tau_o^s + \mathbb{E}_1[\max\{W^*(y) - \frac{\tau_d}{2}y, 0\}]\big| < C|x - y|$$

which means the Lipschitz continuity condition of Proposition 1.2.3 in [115] is satisfied. Then according to Proposition 1.2.3 in [115], $\{\gamma_k\}$ converges to γ^*. □

6.4 Performance Evaluation

We use computer simulation to validate our analysis. Consider 5 sources and 4 relays in our network. Channels from sources to relays experience i.i.d. Rayleigh fading while channels from relays to destinations also experience i.i.d. Rayleigh fading. The channel contention parameters are set as: $p_0 = p_1 = 0.3$, $\delta = 20$ μs, $\tau_{RTS} = \tau_{CTS} = 40$ μs, $\tau_d = 2$ ms.

First consider the scenario that the average received signal-to-noise ratio (SNR) in the first and the second hops are the same. When the average SNR varies from 0.5 to 10, Fig. 6.3 shows the numerically calculated (shown as “analytical” in Fig. 6.3) and simulated (shown as “sim” in Fig. 6.3) system throughput of Case I, Case II with intuitive stopping strategy, and Case II with optimal stopping strategy. It can be seen that the analytical and simulation results match well with each other, which confirms the accuracy of the analysis of our three strategies. Next we perform comparison with alternative strategies. In particular, we consider four alternative strategies.

- *Case-I-no-wait* strategy: A winner source has full CSI, and always transmits (i.e., always stop and does not wait). It is equivalent to a stopping strategy with zero

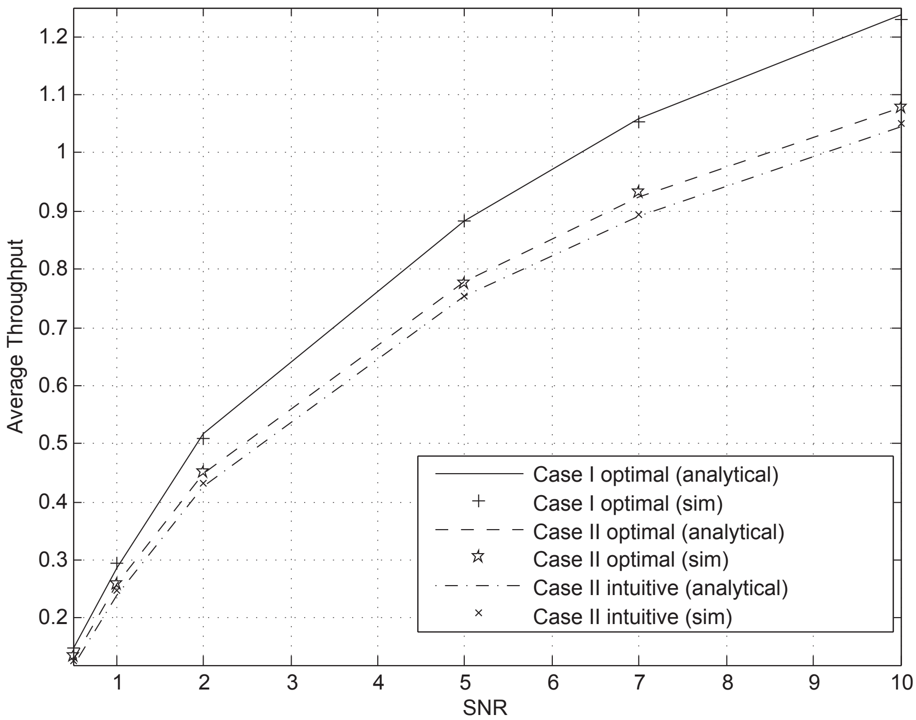

Fig. 6.3. Comparison of analytical and simulation results of our three strategies.

threshold.

- *Case-II-no-wait* strategy: A winner source has partial CSI. And a winner source or relay always transmits. It is equivalent to a bi-layer stopping strategy with zero threshold in the two layers.

- *Case-II-wait-1st-hop* strategy: A winner source has partial CSI. And a winner source applies optimal stopping rule, while a winner relay always transmits. It is equivalent to a bi-layer stopping strategy with zero threshold in the sub-layer.

- *Case-II-wait-2nd-hop* strategy: A winner source has partial CSI. And a winner source always transmits while a winner relay applies optimal stopping rule. It is equivalent to a one-layer stopping strategy in the sub-layer.

Fig. 6.4 shows the average throughput of our three strategies and the four alternative strategies when the first-hop and second-hop SNRs are the same and vary from 0.5 to 10. It can be seen that our optimal strategy in Case I and the Case-I-no-wait strategy have better performance than other strategies. This is because of the full CSI at a winner source. The

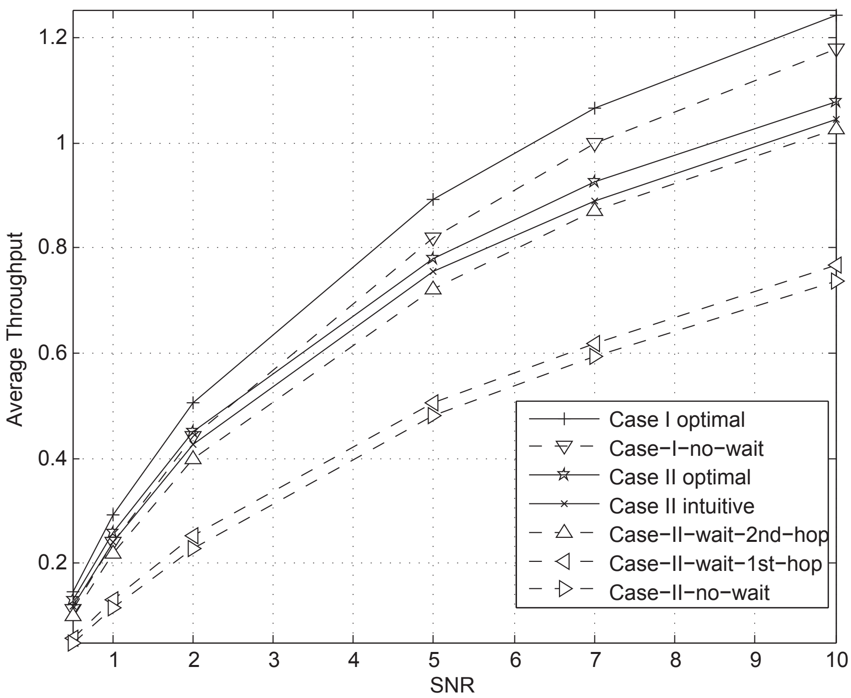

Fig. 6.4. Comparisons of our three strategies with alternative strategies when the first-hop and the second-hop SNRs are the same.

optimal stopping strategy exploits the time diversity of sources (by deciding whether to stop or not) and multi-user diversity of relays (by selecting the best relay). On the other hand, Case-I-no-wait strategy exploits only the multi-user diversity of relays, and therefore, has worse performance than the optimal strategy.

In Case II, among the five strategies, our intuitive strategy and our optimal strategy are the best, with the former having some performance loss compared with the latter, as expected. For the two alternative strategies with a stopping rule applied in one hop, i.e., Case-II-wait-1st-hop and Case-II-wait-2nd-hop, they have a big performance gap, and Case-II-wait-1st-hop strategy is close to the Case-II-no-wait strategy (the worst strategy) while Case-II-wait-2nd-hop strategy is close to our intuitive strategy. The reason is as follows. In Case-II-wait-1st-hop strategy, the threshold in the stopping rule (which is in the first hop) is based on only statistical information of second-hop channels. On the other hand, in Case-II-wait-2nd-hop strategy, the threshold in the stopping rule (which is in the second hop) can be determined based on exact CSI in the first hop. Compared with statistical channel gain information, the exact CSI can help select the *best* threshold.

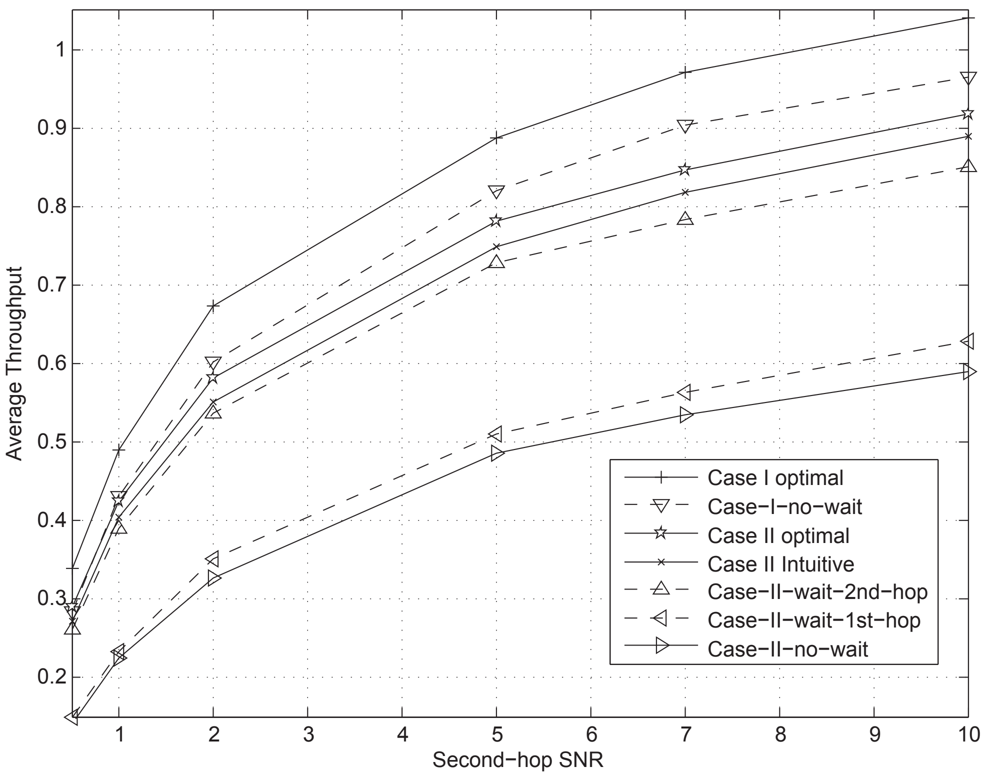

Fig. 6.5. Comparisons of our three strategies with alternative strategies when the first-hop SNR is fixed at 5 and the second-hop SNR varies.

Note that for Case II, our two strategies and the Case-II-wait-2nd-hop strategy have the same communication overhead for a winner source to obtain its CSI with relays and for a winner relay to obtain it CSI with the destination. Since the threshold in the stopping rule of the second hop is dependent on the channel gain realization in the preceding first-hop transmission, a winner source needs to online calculate the threshold (e.g., by iterative algorithms) for the second-hop stopping rule and broadcasts the threshold value together with its packet to all relays. On the other hand, for Case-II-wait-1st-hop strategy, a winner source needs to know its CSI to all relays. But a winner relay does not need its CSI to the destination. And the threshold in stopping rule, which is in the first hop, can be calculated off-line.

Fig. 6.5 shows the comparison of the seven strategies when the first-hop SNR is fixed at 5 and the second-hop SNR varies from 0.5 to 10. Similar observations to those of Fig. 6.4 can be obtained.

6.5 Conclusion

In a wireless relay network, the sources and relays all experience independent fading. It is desired to exploit the multi-source diversity, multi-relay diversity, and time diversity. To achieve this, OCA is needed, which is investigated in our research in a distributed structure. For the two considered cases (with a winner source having or not having CSI of the second hop), we derive optimal stopping strategies for OCA. This research should provide insights to the design of channel-aware MAC protocols in wireless relay network. Further research may include the cases with quantized CSI and with quality-of-service constraints.

Appendix 6.A: Proof of Existence of Optimal Stopping Rule (6.25)

To guarantee that for $\gamma \geq 0$ an optimal rule N^* which achieves $V^*(\gamma)$ exists, two conditions should be satisfied:

$$\limsup_{n\to\infty}\{R_n^1 - \gamma(R_n^2 + \frac{\tau_d}{2} + \sum_{l=1}^{n} t_l^s)\} = -\infty. \tag{6.44}$$

$$\mathbb{E}_1\big[\sup_n\{R_n^1 - \gamma(R_n^2 + \frac{\tau_d}{2} + \sum_{l=1}^{n} t_l^s)\}\big] < \infty. \tag{6.45}$$

By decomposition of left hand side of (6.45), we have

$$\begin{aligned}\mathbb{E}_1\big[\sup_n\{R_n^1 - \gamma(R_n^2 + \frac{\tau_d}{2} + \sum_{l=1}^{n} t_l^s)\}\big] \leq \mathbb{E}_1\big[\sup_n\{R_n^1 - n\gamma(\tau_o^s - \varepsilon)\}\big] + \\ \mathbb{E}_1\big[\sup_n\{\gamma\sum_{l=1}^{n}(\tau_o^s - \varepsilon - t_l^s)\}\big] - \gamma\frac{\tau_d}{2} + \mathbb{E}_1\big[\sup_n\{-\gamma R_n^2\}\big]\end{aligned} \tag{6.46}$$

where $0 < \varepsilon < \tau_o^s$.

Similar to proofs of Lemmas 6.1 and 6.2, if we prove $\mathbb{E}[(R_n^1)^2] < \infty$ and $\mathbb{E}[(R_n^2)^2] < \infty$, the existence conditions (6.44) and (6.45) are satisfied. Therefore, next we focus on proofs of $\mathbb{E}[(R_n^1)^2] < \infty$ and $\mathbb{E}[(R_n^2)^2] < \infty$. We consider proof of $\mathbb{E}[(R_n^2)^2] < \infty$ first.

From Corollary 6.1, we have $R_n^1 = \frac{\lambda^*\tau_o^r}{1-F_{R_m}(\lambda^*)} + \frac{\lambda^*\tau_d}{2}$, $R_n^2 = \frac{\tau_o^r}{1-F_{R_m}(\lambda^*)} + \frac{\tau_d}{2}$, and

$$\mathbb{E}_1[(R_n^2)^2] = (\tau_o^r)^2 \cdot \mathbb{E}_1\big[\big(\frac{1}{1-F_{R_m}(\lambda^*)}\big)^2\big] + \tau_o^r\tau_d \cdot \mathbb{E}_1[\frac{1}{1-F_{R_m}(\lambda^*)}] + \frac{\tau_d^2}{4}. \tag{6.47}$$

If $\mathbb{E}_1\big[\big(\frac{1}{1-F_{R_m}(\lambda^*)}\big)^2\big] < \infty$, we have $\mathbb{E}_1[\frac{1}{1-F_{R_m}(\lambda^*)}] < \infty$, and then $\mathbb{E}_1[(R_n^2)^2] < \infty$ is proved. Thus, we calculate $1 - F_{R_m}(\lambda^*)$ and prove $\mathbb{E}_1\big[\big(\frac{1}{1-F_{R_m}(\lambda^*)}\big)^2\big] < \infty$.

Conditioned on $\mathcal{F}$, we have

$$
\begin{aligned}
1 - F_{R_m}(\lambda^*) &= \mathbb{P}\big(\sum_{j=1}^{L} \mathbb{I}[s(m) = j] \log_2(1 + \frac{P_s P_r |f_{s(n)j}(n)|^2 |g_{js(n)}(m)|^2}{1 + P_s|f_{s(n)j}(n)|^2 + P_r|g_{js(n)}(m)|^2}) \geq \lambda^*\big) \\
&= \frac{1}{L}\sum_{j=1}^{L} \mathbb{P}\Big(\log_2(1 + \frac{P_s P_r |f_{s(n)j}(n)|^2 |g_{js(n)}(m)|^2}{1 + P_s|f_{s(n)j}(n)|^2 + P_r|g_{js(n)}(m)|^2}) \geq \lambda^*\Big) \\
&= \frac{1}{L}\sum_{j=1}^{L} \mathbb{P}\Big(P_r|g_{js(n)}(m)|^2\big(P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1)\big) \geq (2^{\lambda^*}-1)\left(1 + P_s|f_{s(n)j}(n)|^2\right)\Big) \\
&= \frac{1}{L}\sum_{j=1}^{L} \mathbb{P}\Big(P_r|g_{js(n)}(m)|^2 \geq \frac{(2^{\lambda^*}-1)(1+P_s|f_{s(n)j}(n)|^2)}{P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1)}\Big)\mathbb{I}\big[P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1) > 0\big] \\
&\quad + \frac{1}{L}\sum_{j=1}^{L} \mathbb{P}\Big(P_r|g_{js(n)}(m)|^2 < \frac{(2^{\lambda^*}-1)(1+P_s|f_{s(n)j}(n)|^2)}{P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1)}\Big)\mathbb{I}\big[P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1) < 0\big] \\
&= \frac{1}{L}\sum_{j=1}^{L} e^{-\frac{1}{\sigma_g^2 P_r}\frac{(2^{\lambda^*}-1)(1+P_s|f_{s(n)j}(n)|^2)}{P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1)}}\mathbb{I}\big[P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1) > 0\big]
\end{aligned}
$$

which leads to

$$
\mathbb{E}_1\big[(\frac{1}{1 - F_{R_m}(\lambda^*)})^2\big] \leq L^2 \cdot \mathbb{E}_1[\Lambda] \tag{6.48}
$$

where $\Lambda \triangleq 1/\Big(\sum_{j=1}^{L} e^{-\frac{2}{\sigma_g^2 P_r}\frac{(2^{\lambda^*}-1)(1+P_s|f_{s(n)j}(n)|^2)}{P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1)}}\mathbb{I}\big[P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1) > 0\big]\Big)$.

So, if $\mathbb{E}_1[\Lambda]$ is finite, we should have $\mathbb{E}_1\big[\big(\frac{1}{1-F_{R_m}(\lambda^*)}\big)^2\big] < \infty$. And $\mathbb{E}_1[\Lambda]$ is proved to be finite, if Λ is finite when $|f_{s(n)j}(n)|^2 \to 0,\ \forall j$ and when $|f_{s(n)j}(n)|^2 \to \infty,\ \forall j$, as follows.

It is proved, by using proof by contradiction, that partial derivatives $\lim_{P_s|f_{s(n)j}(n)|^2 \to 0, \forall j} \frac{\partial(2^{\lambda^*}-1)}{\partial P_s|f_{s(n)i}(n)|^2}$ are not equal to 1's.

Then we have

$$
\begin{aligned}
\lim_{P_s|f_{s(n)j}(n)|^2 \to 0, \forall j} \Lambda &= \lim_{P_s|f_{s(n)j}(n)|^2 \to 0, \forall j} 1/\Big\{\sum_{j=1}^{L} e^{-\frac{2}{\sigma_g^2 P_r}\frac{(2^{\lambda^*}-1)(1+P_s|f_{s(n)j}(n)|^2)}{P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1)}}\Big\} \\
&= 1/\Big\{\sum_{j=1}^{L} \lim_{P_s|f_{s(n)j}(n)|^2 \to 0, \forall j} e^{-\frac{2}{\sigma_g^2 P_r}\frac{(2^{\lambda^*}-1)(1+P_s|f_{s(n)j}(n)|^2)}{P_s|f_{s(n)j}(n)|^2 - (2^{\lambda^*}-1)}}\Big\} \\
&= 1/\Big\{\sum_{j=1}^{L} e^{-\frac{2}{\sigma_g^2 P_r}\frac{1}{\lim_{P_s|f_{s(n)j}(n)|^2 \to 0, \forall j} \frac{\partial(2^{\lambda^*}-1)}{\partial P_s|f_{s(n)j}(n)|^2}\big|_{P_s|f_{s(n)j}(n)|^2=0} - 1}}\Big\} < \infty.
\end{aligned}
$$

From (6.23) and using similar derivation as (6.19), we have $\lambda^* \leq \frac{\tau_d}{2\tau_o^r} \sum_{j=1}^{L} \frac{1}{L} \frac{1}{\ln 2} P_r \sigma_g^2$, which leads to

$$\lim_{P_s |f_{s(n)j}(n)|^2 \to \infty, \forall j} \Lambda = 1/\{\sum_{j=1}^{L} e^{-\frac{2}{\sigma_g^2 P_r}}\} < \infty. \tag{6.49}$$

By (6.48), $\mathbb{E}_1[(\frac{1}{1-F_{R_m}(\lambda^*)})^2] < \infty$ is proved, and $\mathbb{E}_1[(R_n^2)^2] < \infty$.

Since $\lambda^* \leq \frac{\tau_d}{2\tau_o^r} \sum_{j=1}^{L} \frac{1}{L} \frac{1}{\ln 2} P_r \sigma_g^2$, we have

$$\mathbb{E}_1[(R_n^1)^2] = \mathbb{E}_1[\lambda^{*2}(R_n^2)^2] \leq \mathbb{E}_1[(\frac{\tau_d}{2\tau_o^r} \sum_{j=1}^{L} \frac{1}{L} \frac{1}{\ln 2} P_r \sigma_g^2)^2 (R_n^2)^2] < \infty.$$

Chapter 7

Conclusions and Future Research Works

This chapter summarizes the contributions of the thesis work, and discusses future works.

7.1 Conclusions

Being the main focus of this thesis, two major topics are studied. The first topic is related to OCSA in CRNs. The cases with and without statistical information of primary traffic are analyzed. Optimal (or order optimal) strategies are derived. The second topic is related to OCA in distritbuted wireless cooperative networks. Optimal solutions are derived theoretically. The effectiveness and efficiency of the derived strategies in the two topics are verified numerically.

In Chapter 3, when statistic information of primary traffic is known and sensing is imperfect, the optimal strategies that decide on the optimal set of channels to sense and the optimal set of sensed-free channels to access are derived. Interestingly, in general, when the secondary user can access only a limited number of channels, it may not be optimal to sense the channels with the largest rewards.

In Chapter 4, when the busy/idle states of each channel follow the i.i.d. model, OCSA strategies are derived. If the secondary user can sense all potential channels simultaneously, OCSA strategies with asymptotically finite regrets are derived. If the secondary user can simultaneously sense only a subset of the potential channels, it is impossible for any OCSA strategies to have finite regrets, and the best possible strategies are those with asymptotically logarithmic regrets, which are derived in Chapter 4. Further, in Chapter 5, when the busy/idle states of each channel follow the Markov model, an OCSA strategy with asymptotically logarithmic regrets is derived. The low complexity of the strategies in these two chapters can largely facilitate implementation of those strategies in real CRNs in the near future.

In Chapter 6, optimal OCA in a distributed AF relay network is investigated. In two considered cases depending on whether a winner source has or does not have CSI of the second hop transmission, optimal stopping strategies are derived that maximize the average system throughput. In the proposed strategies, the first-hop stopping rule is pure-threshold, and the second-hop stopping rule (for the case that a winner source does not have CSI of the second hop) has a threshold determined by channel gain realization in the preceding first-hop transmission. The easy implementation of the strategies is of interest in potential applications.

7.2 Future Research Directions

In OCSA in CRNs, competition among multiple secondary users is worth further investigation. In a distributed CRN, multiple secondary users sense and access the primary channels independently, and thus, collision between secondary transmissions is inevitable. Under such circumstance, the problem of how to design an efficient multiple-user OCSA rule is much more challenging than the single-user case, and will be investigated.

In Chapter 6, optimal distributed OCA in wireless cooperative networks is investigated. As multiple relays are deployed, a winner source has to probe the channels from itself to all relays. This process demands large information overhead, and the system efficiency is affected. Then the following question arises: is it necessary to probe all the relays? In other words, after the winner source gets CSI to some relays, the winner source may stop probing other relays and use only those probed relays. Then the optimal stopping strategy that decides when to stop probing relays is of interest, which should be jointly designed with the winner source's strategy (on whether to give up transmission opportunity or utilize the channel) and the relays' stopping strategies.

In the OCA for wireless cooperative networks in this thesis, after channel probing, transmitters get perfect CSI from receivers, and use adaptive modulation according to the channel conditions. Since perfect CSI is difficult to obtain in practice, limited CSI feedback is a widely accepted case. The limited CSI feedback motivates a new research problem. In such problem, how to design an optimal code-book for feedback information and how to design optimal OCA are of interest, and should be jointly investigated.

References

[1] FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group," Nov. 2002.

[2] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, Royal Institute of Technology, Stockholm, Sweden, 2000.

[3] C. Cordeiro, K. Challapali, D. Birru, and N. Sai, "IEEE 802.22: An introduction to the first wireless standard based on cognitive radios," *Journal of Communications*, vol. 1, no. 1, pp. 38–47, Apr. 2006.

[4] C. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. Shellhammer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Communications Magzine*, vol. 47, no. 1, pp. 130– 138, Jan. 2009.

[5] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[6] R. Zhang, Y.-C. Liang, and S. Cui, "Dynamic resource allocation in cognitive radio networks," *IEEE Signal Processing Magzine*, vol. 27, no. 3, pp. 102–114, May 2010.

[7] H. Jiang, L. Lai, R. Fan, and H. Poor,, "Optimal selection of channel sensing order in cognitive radio," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 297–307, Jan. 2009.

[8] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput trade-off for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.

[9] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, “Breaking spectrum gridlock with cognitve radios: an information theoretic perspective,” *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[10] A. F. Molisch, L. J. Greenstein, and M. Shafi, “Propagation issues for cognitive radio,” *Proceedings of the IEEE*, vol. 97, no. 5, pp. 787–804, May 2009.

[11] R. Fan and H. Jiang, “Optimal multi-channel cooperative sensing in cognitive radio networks,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, pp. 1128–1138, Mar. 2010.

[12] L. Lai, H. El Gamal, H. Jiang, and H. V. Poor, “Cognitive medium access: Exploration, exploitation, and competition,” *IEEE Transactions on Mobile Computing*, vol. 10, no. 2, pp. 239–253, Feb. 2011.

[13] R. Fan, H. Jiang, Q. Guo, and Z. Zhang, “Joint optimal cooperative sensing and resource allocation in multichannel cognitive radio networks,” *IEEE Transactions on Vehicular Technology*, vol. 60, no. 2, pp. 722–729, Feb. 2011.

[14] R. Zhang, S. Cui, and Y.-C. Liang, “On ergodic sum capacity of fading cognitive multiple-access and broadcast channels,” *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5161–5178, Nov. 2009.

[15] Z. Quan, S. Cui, A.H. Sayed, and H.V. Poor, “Optimal multiband joint detection for spectrum sensing in cognitive radio networks,” *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.

[16] J. Jin and B. Li, “Cooperative resource management in cognitive WiMAX with femto cells,” in *Proceedings of IEEE INFOCOM 2010*.

[17] R. Urgaonkar and M. J. Neely, “Opportunistic scheduling with reliability guarantees in cognitive radio networks,” *IEEE Transactions on Mobile Computing*, vol. 8, no. 6, pp. 766–777, June 2009.

[18] S. Atapattu, C. Tellambura, and H. Jiang, “Energy detection based cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1232–1241, Apr. 2011.

[19] H.-S. Chen, W. Gao, and D. Daut, "Signature based spectrum sensing algorithms for IEEE 802.22 WRAN," in *Proceedings of IEEE International Conference on Communications (ICC)*, 2007, pp. 6487–6492.

[20] W. Gardner, "Signal interception: a unifying theoretical framework for feature detection," *IEEE Transactions on Communications*, vol. 36, no. 8, pp. 897–906, Aug. 1988.

[21] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.

[22] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity–Part I: System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[23] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity–Part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[24] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[25] A. Scaglione and Y.-W. Hong, "Opportunistic large arrays: Cooperative transmission in wireless multi-hop ad hoc networks for the reach back channel," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2082–2092, Aug. 2003.

[26] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[27] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magzine*, vol. 39, no. 2, pp. 150–154, Feb. 2001.

[28] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[29] S. Gao, L. Qian, D.R. Vaman, and Z. Han, "Distributed cognitive sensing for time varying channels: exploration and exploitation," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC) 2010*.

[30] X. Fang, D. Yang, and G. Xue, "Taming wheel of fortune in the air: An algorithmic framework for channel selection strategy in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 2, pp. 783–796, Feb. 2013.

[31] A. Anandkumar, N. Michael, A.K. Tang, and A. Swami, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 731–745, Apr. 2011.

[32] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[33] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad hoc networks with random access: an optimal stopping approach," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 205–222, Jan. 2009.

[34] W. Ge, J. Zhang, J. Wieselthier, and X. Shen, "PHY-aware distributed scheduling for ad hoc communications with physical interference model," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2682–2693, May 2009.

[35] S. Tan, D. Zheng, J. Zhang, and J. Zeidler, "Distributed opportunistic scheduling for ad-hoc communications under delay constraints," in *Proceedings of IEEE INFOCOM 2010*.

[36] T. Ferguson, *Optimal stopping and applications*, Available online: http://www.math.ucla.edu/~tom/Stopping/Contents.html, Mathematics Department, UCLA.

[37] N. Devroye, P. Mitran, and V. Tarokh, "Cognitive decomposition of wireless networks," in *Proceedings of CrownCom 2006*, Mykonos, Greece, June 2006.

[38] IEEE 802.11 WG, IEEE 802.11e/D11, IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and

metropolitan area networks-Specific requirements-Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 7: Medium Access Control (MAC) Quality of Service (QoS) Enhancements, Oct. 2004.

[39] I. F. Akyildiz, D. A. Levine, and I. Joe, “A slotted CDMA protocol with BER scheduling for wireless multimedia networks," *IEEE/ACM Transactions on Networking*, vol. 7, no. 2, pp. 146–158, Apr. 1999.

[40] V. Huang and W. Zhuang, “QoS-oriented packet scheduling for wireless multimedia CDMA communications," *IEEE Transactions on Mobile Computing*, vol. 3, no. 1, pp. 73–85, Jan-Feb. 2004.

[41] N. Abramson, “The ALOHA system - another alternative for computer communications," in *Proceedings of 1970 AFIPS Fall Joint Computer Conference*, pp. 281–285, 1970.

[42] L. G. Roberts, “ALOHA packet system with and without slots and capture," *ACM SIGCOMM Computer Communication Review*, vol. 5, no. 2, pp. 28–42, Apr. 1975.

[43] D. Shen and V. O. K. Li, “Stabilized multi-channel ALOHA for wireless OFDM networks,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 701–705, 2002.

[44] Y.-J. Choi, S. Park, and S. Bahk, “Multichannel random access in OFDMA wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 603–613, Mar. 2006.

[45] IEEE 802.11 WG, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, Standard, IEEE, Aug. 1999.

[46] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, “Multiuser OFDM with adaptive subcarrier, bit and power allocation,” *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1757, Oct. 1999.

[47] K. Kim, Y. Han, and S.-L. Kim, “Joint subcarrier and power allocation in uplink OFDMA systems,” *IEEE Communications Letters*, vol. 9, no. 6, pp. 526–528, June 2005.

[48] L. Gao and S. Cui, "Efficient subcarrier, power, and rate allocation with fairness consideration for OFDMA uplink," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1507–1511, May 2008.

[49] C. Y. Ng and C. W. Sung, "Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1667–1675, May 2008.

[50] J. Huang, V. G. Subramanian, R. Agrawal, and R. Berry, "Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 226–234, Feb. 2009.

[51] G. Song, Y. Li, L. J. Cimini, and H. Zheng, "Joint channel-aware and queue-aware data scheduling in multiple shared wireless channels," in *Proceedings of IEEE Wireless Communication and Networking Conference (WCNC) 2004*, pp. 1939–1944, Mar. 2004.

[52] A. M. El-Hajj, E. Yaacoub, and Z. Dawy, "Uplink OFDMA resource allocation with ergodic rate maximization," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2009*, pp. 276–280, Sep. 2009.

[53] E. Yaacoub, A. M. El-Hajj, and Z. Dawy, "Weighted ergodic sum-rate maximisation in uplink orthogonal frequency division multiple access and its achievable rate region," *IET Communications*, vol. 4, no. 18, pp. 2217–2229, 2010.

[54] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 3, pp. 92–106, 2009.

[55] D. Wu, D. Yu, and Y. Cai, "Subcarrier and power allocation in uplink OFDMA systems based on game theory," in *Proceedings of IEEE International Conference on Neural Networks and Signal Processing*, Zhenjiang, China, June 2008.

[56] M. Kobayashi and G. Caire, “Iterative waterfilling for weighted rate sum maximization in MIMO-MAC,” in *Proceedings of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC) 2006*, Cannes, France, July 2006.

[57] N. Ul Hassan and M. Assaad, “Low complexity margin adaptive resource allocation in downlink MIMO-OFDMA system," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3365–3371, July 2009.

[58] G. Song and Y. Li, “Cross-layer optimization for OFDM wireless networks-Part I: theoretical framework,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, Mar. 2005.

[59] G. Song and Y. Li, “Cross-layer optimization for OFDM wireless networks-Part II: algorithm development,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[60] I. C. Wong and B. L. Evans, “Optimal downlink OFDMA resource allocation with linear complexity to maximize ergodic rates,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 962–971, Mar. 2008.

[61] P. S. Chow, J. C. Tu, and J. M. Cioffi, “A discrete multitone transceiver system for HDSL applications,” *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 6, pp. 895–908, Aug. 2011.

[62] A. L. Stolyar, “Maximizing queueing network utility subject to stability: greedy primal-dual algorithm,” *Queueing Systems*, vol. 50, no. 4, pp. 401–457, 2005.

[63] L. Li and A. J. Goldsmith, “Optimal resource allocation for fading broadcast channels-Part I: ergodic capacity,” *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[64] A. J. Goldsmith and M. Effros, “A discrete multitone transceiver system for HDSL applications,” *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 6, pp. 895–908, Aug. 2011.

[65] W. Yu and J. M. Cioffi, “FDMA capacity of gaussian multiple-access channels with ISI,” *IEEE Transactions on Communications*, vol. 50, no. 1, pp. 102–111, Jan. 2002.

[66] L. M. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, “Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms,” *IEEE Transactions on Communications*, vol. 52, no. 6, pp. 922–930, June 2004.

[67] X. Wang and G. B. Giannakis, “Ergodic capacity and average rate-guaranteed scheduling for wireless multiuser OFDM systems,” in *Proceedings of IEEE International Symposium on Information Theory (ISIT) 2008*, pp. 1691–1695, July 2008.

[68] A. Ozgur, O. Leveque, and D. Tse, “Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks,” *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3549–3572, Oct. 2007.

[69] Q. Li, K. H. Li, and K. C. Teh, “An achievable rate region for the cognitive interference channel with causal bi-directional cooperation,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1721–1728, May 2010.

[70] Q. Li, K. H. Li, and K. C. Teh, “Diversity-multiplexing trade-Off of fading interference channels with source cooperation and partial CSIT,” *IEEE Transactions on Information Theory*, vol. 57, pp. 2781–2804, May 2011.

[71] V. R. Cadambe and S. A. Jafar, “Interference alignment and degrees of freedom of the K-user interference channel,” *IEEE Transactions on Information Theory*, vol. 54, pp. 3425–3441, Aug. 2008.

[72] H. Zhu and G. Cao, “On improving the performance of IEEE 802.11 with relay-enabled PCF,” *Springer Journal on Mobile Networks and Applications*, vol. 54, pp. 423–434, Nov. 2004.

[73] H. Zhu and G. Cao, “rDCF: A relay-enabled medium access control protocol for wireless ad hoc networks,” *IEEE Transactions on Mobile Computing*, vol. 5, no. 9, pp. 1201–1214, Sep. 2006.

[74] G. R. Hiertz, S. Max, R. Zhao, D. Denteneer, and L. Berlemann, “Principles of IEEE 802.11s,” in *Proceedings of International Conference on Computer Communications and Networks (ICCCN)*, pp. 1002–1007, Aug. 2007.

[75] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and A. Brunstrom, "Towards systems beyond 3G based on adaptive OFDMA transmission," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2432–2455, Dec. 2007.

[76] IEEE P802.16j/D9, "Draft IEEE standard for local and metropolitan area networks Part 16: Air interface for fixed and mobile broadband wireless access systems: multihop relay specification," Feb. 2009.

[77] Y.-W. Hong and A. Scaglione, "Energy-efficient broadcasting with cooperative transmission in wireless ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 10, pp. 2844–2855, Oct. 2006.

[78] M. C. Vuran and I.F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, pp. 316–329, Apr. 2006.

[79] Y.-W. Hong and A. Scaglione, "On multiple access for correlated sources: A content-based group testing approach," in *Proceedings of IEEE Information Theory Workshop*, pp. 298–303, Oct. 2004.

[80] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 613–626, Feb. 2006.

[81] S. Huang, X. Liu, and Z. Ding, "Optimal transmission strategies for dynamic spectrum access in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 12, pp. 1636–1648, Dec. 2009.

[82] R. Fan and H. Jiang, "Channel sensing-order setting in cognitive radio networks: A two-user case," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 4997–5008, Nov. 2009.

[83] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magzine*, vol. 46, no. 4, pp. 40–48, Apr. 2008.

[84] A. K. L. Yau, P. Komisarczuk, and P. D. Teal, "On multi-channel MAC protocols in cognitive radio networks," in *Proceedings of Australasian Telecommunication Networks and Applications Conference (ATNAC)*, pp. 300–305, Dec. 2008.

[85] H. Wang, H. Qin, and L. Zhu, "A survey on MAC protocols for opportunistic spectrum access in cognitive radio networks," in *Proceedings of International Conference on Computer Science and Software Engineering*, pp. 214-218, Dec. 2008.

[86] P. Pawelczak, S. Pollin, H. S. W. So, A. Bahai, R. V. Prasad, and R. Hekmat, "Performance analysis of multichannel medium access control algorithms for opportunistic spectrum access," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 3014–3031, July 2009.

[87] T. L. Lai, and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, no. 1 pp. 4–22, 1985.

[88] V. Anantharam, P. Varaiya, and J. Walrand, "Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays-Part I: I.I.D. rewards," *IEEE Transactions on Automatic Control*, vol. 32, no. 11, pp. 968–976, Nov. 1987.

[89] R. Agrawal, "Sample mean based index policies with O(log n) regret for the multi-armed bandit problem," *Advances in Applied Probability*, vol 27, no. 4, pp. 1054–1078, Dec. 1995.

[90] H. T. Cheng and W. Zhuang, "Simple channel sensing order in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 676–688, Apr. 2011

[91] Y. Pei, Y.-C. Liang, K. C. Teh, and K. H. Li, "Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1648–1659, Sept. 2011.

[92] Y. Pei, Y.-C. Liang, K. C. Teh, and K. H. Li, "How much time is needed for wideband spectrum sensing?" *IEEE Transactions on Wireless Communications*, vol. 8, no. 11, pp. 5466–5471, Nov. 2009.

[93] T. V. Nguyen, H. Shin, T. Q. S. Quek, and M. Z. Win, "Sensing and probing cardinalities for active cognitive radios," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1833–1848, Apr. 2012.

[94] M.J. Neely, “Opportunistic scheduling with worst case delay guarantees in single and multi-hop networks,” in *Proceedings of IEEE INFOCOM 2011*, pp. 1728–1736.

[95] C.-C. Wang, S.R. Kulkarni, and H.V. Poor, “Bandit problems with side observations,” *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 338–355, Mar. 2005.

[96] A. Dembo and O. Zeitouni, Large deviation techniques and applications. New York, NY: Springer-Verlag, 1998.

[97] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multiarmed bandit problem,” *Machine Learing*, vol. 47, no. 2-3, pp. 235–256, May-June 2002.

[98] L. Kocsis and C. Szepesvari, “Bandit based Monte-Carlo planning,” in *Proceedings of European Conference on Machine Learning*, pp. 282–293, 2006.

[99] Y.S. Chow and H. Robbins, and D. Siegmund, Great expectations: The theory of optimal stopping. Boston, MA: Houghton Mifflin, 1971.

[100] C. Tekin and M. Liu, “Online learning in opportunistic spectrum access: a restless bandit approach,” in *Proceedings of IEEE INFOCOM 2011*, pp. 2462–2470.

[101] V. Anantharam, P. Varaiya, and J. Walrand, “Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays-Part II: Markovian rewards,” *IEEE Transactions on Automatic Control*, vol. 32, no. 11, pp. 977–982, Nov.1987.

[102] C. Tekin and M. Liu, “Online learning of rested and restless bandits,” *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5588–5611, Aug. 2012.

[103] P. Lezaud, “Chernoff-type bound for finite Markov chains," *The Annals of Applied Probability*, vol. 8, no. 3, pp. 849-867, Aug. 1998.

[104] L.-L. Xie and P. R. Kumar, “An achievable rate for the multiple-level relay channel." *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.

[105] Y. Jing and H. Jafarkhani, “Single and multiple relay selection schemes and their achievable diversity orders,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1414–1423, Mar. 2009.

[106] J. Zhang, J. Jia, Q. Zhang, and E. M. K. Lo, “Implementation and evaluation of cooperative communication schemes in software-defined radio testbed," in *Proceedings of IEEE INFOCOM 2010*.

[107] Q. Zhang, J. Jia, and J. Zhang, “Cooperative relay to improve diversity in cognitive radio networks," *IEEE Communications Magzine*, vol. 47, no. 2, pp. 111-117, Feb. 2009.

[108] D. Gunduz and E. Erkip, “Opportunistic cooperation by dynamic resource allocation,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1446–1454, Apr. 2007.

[109] Y. Jing and H. Jafarkhani, “Network beamforming using relays with perfect channel information,” *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2499–2517, June 2009.

[110] Y. Li, B. Vucetic, Z. Zhou, and M. Dohler, “Distributed adaptive power allocation for wireless relay networks,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 948–958, Mar. 2007.

[111] A. Sadek, W. Su, and K. Liu, “Multinode cooperative communications in wireless networks,” *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 341–355, Jan. 2007.

[112] X. Tang and Y. Hua, “Optimal design of non-regenerative MIMO wireless relays,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.

[113] N. Khajehnouri and A. Sayed, “Distributed MMSE relay strategies for wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3336–3348, July 2007.

[114] Y. Fan and J. Thompson, “MIMO configurations for relay channels: Theory and practice,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1774–1786, May 2007.

[115] D.P. Bertsekas, Nonlinear Programming. Athena scientific, 1999.