Computer Networks 91 (2015) 117–134

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

A Chinese restaurant game for learning and decision making in cognitive radio networks☆

Biling Zhang [a,b,*], Yan Chen [c], Chih-Yu Wang [d], K.J. Ray Liu [c]

[a] School of Network Education, Beijing University of Posts and Telecommunications, Beijing 100876, PR China
[b] The State Key Laboratory of Integrated Services Networks, Xidian University, Xian 710071, PR China
[c] Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA
[d] Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 5 September 2014
Revised 28 July 2015
Accepted 21 August 2015
Available online 1 September 2015

Keywords:
Chinese restaurant game
Opportunistic spectrum access
Cooperative sensing
Network externality
Game theory
Social learning

ABSTRACT

In cognitive radio networks, secondary users (SUs) are allowed to opportunistically exploit the licensed channels. Once finding the spectrum holes, SUs need to share the available licensed channels. Therefore, one of the critical challenges for fully utilizing the spectrum resources is how the SUs obtain accurate information about the primary users' (PUs') activities and make right decisions on which channels to access so as to avoid competition from other SUs. In this paper, we formulate SUs' learning and decision making process as a Chinese restaurant game, which is concerned with negative network externality, by considering the scenario where each SU senses only one of the channels and then makes access decisions sequentially. In the proposed game, SUs build the knowledge of the PUs' activities by their own sensing and learning the information from other SUs. They also predict the subsequent SUs' decisions to maximize their own utilities. We analyze the interactions among SUs and study specifically the impact of SUs' initial belief, sensing accuracy and channel quality on their decisions. We also derive the theoretical results for the two-user two-channel case and extend the results to the multi-user multi-channel case. Finally, we verify the theoretical results and evaluate the performance of the proposed scheme in terms of social welfare through simulations.

1. Introduction

During the last decades, the demand for wireless spectrum resources has grown rapidly. Cognitive radio technology as a promising way to increase the efficiency of spectrum utilization and thus alleviate the spectrum shortage has drawn great attentions. In cognitive radio networks, secondary users (SUs) as unlicensed users are allowed to use licensed spectrum bands with the constraint that they do not incur harmful interference to the primary users (PUs) who own the license of the spectrum bands. One typical cognitive radio technology is opportunistic spectrum access, where SUs need to perform spectrum sensing, i.e., detect the PUs' activities, and access the spectrum once finding spectrum holes.

In the literature, spectrum sensing approaches are proposed to identify spectrum holes [2,3], while spectrum access schemes aim at designing medium access control protocols to efficiently share the available spectrum resources among SUs [4,5]. Joint spectrum sensing and access are also considered in works such as [6], [7] and [8]. Although existing dynamic spectrum access schemes have improved the spectrum utilization efficiency, due to the mobility of nodes and the dynamics of the channel variation, the accuracy of users' decisions is limited and remains a challenge to fully utilize the scarce spectrum resources. The analysis of

☆ Part of this work have been presented at the 2012 IEEE Global Telecommunications Conference (Globecom)[1].

* Corresponding author. Tel.: 86-13691581992.
E-mail addresses: bilingzhang@bupt.edu.cn (B. Zhang), yan@umd.edu (Y. Chen), tomkywang@gmail.com (C.-Y. Wang), kjrliu@umd.edu (K.J. Ray Liu).

http://dx.doi.org/10.1016/j.comnet.2015.08.010
1389-1286/

fine-grain spectrum usage traces in [9] showed that even with extensive statistical knowledge on PUs' access patterns, SUs can only extract 20–30% of available spectrum due to improper access strategies.

SUs' inaccurate decision making mainly results from the lack of knowledge on the state of licensed channels. Fortunately, SUs' limited knowledge can be expanded through learning. An SU can learn from its previous experience [10]. In the context of cognitive radio network, machine learning technology such as Q-learning [11] has been well investigated. Alternatively, an SU can learn from some other information sources, such as signals revealed by neighboring SUs or the actions they have taken. Such a type of learning is referred to as social learning, where cooperative spectrum sensing is regarded as one of the typical applications. By jointly combining sensing information from SUs over a wide area, spatial diversity among these SUs is exploited and the reliability of spectrum sensing is improved [12,13].

Most of existing works [10,14] study how users' believes are formed through learning and how accurate the believes will be when more information is collected, assuming that there is no network externality, i.e., a user's reward will not be affected by the actions of other users. However, such an assumption is generally not true. In reality, not only channel availability but also channel quality influences an SU's utility obtained from accessing a channel. Therefore, it is very likely that two or more SUs will choose the same channel, especially when the number of SUs is much larger than the number of available channels. An SU then usually needs to share the channel with others, but due to negative network externality, the more SUs select the same channel, the less utility each SU can obtain from this channel. In such a case, when a rational but selfish SU chooses the best channel from multiple available channels to maximize its utility, it must take into account other SUs' decisions, which leads to a game among SUs. In the literature, the interactions of SUs are usually formulated as a global game [15] or a congestion game [16], where all users make decisions simultaneously and a user's reward is determined by the system state and the number of users making the same decision with him/her.

While lots of works have been done regarding the simultaneous decision making problem, the sequential decision making problem has not been well investigated mainly due to the complexity of the problem. Unlike the simultaneous decision problem where all SUs make access decisions and announce them at the same time and thereby no information is exchanged among SUs, in sequential decision making problem where SUs make and announce decisions one by one, the signals and decisions revealed by the previous SUs can be overheard by the following SUs. An SU therefore can exploit not only the signal observed by its own sensing, but also information from other SUs, which complicates the problem but helps to enhance the reliability of SUs' estimation on spectrum occupancy. On the other hand, besides taking previous SUs' actions into consideration, an SU's optimal strategy in sequential decision making problem should involve predicting decisions of subsequent SUs. Due to the fact that information collected by different SUs is asymmetric, i.e., an SU making decision later can collect more information, such a prediction is more complicated than that in a simultaneous decision case. However, in practice, sequential decision making scenarios are even more prevailing, especially in cognitive radio networks where SUs do not know each other and synchronization among them is quite difficult.

Therefore, in this paper, we focus on the problem of distributed channel selection under the scenario where SUs have to make decisions sequentially after each of them senses only one of the channels, and tries to provide some insights and results on it. The learning and decision making process of SUs is modeled as a Chinese restaurant game which provides a general framework for social learning problems with negative network externality.

To illustrate a typical Chinese restaurant game, consider a Chinese restaurant with K tables and M rational customers. Assume each customer can request for one table. The table requested by the ith customer is x_i, and the size of the ith table is T_i but unknown to customers. Since customers prefer bigger dining space, they may prefer bigger tables. However, a customer may need to share the table with others if multiple customers request for the same table. Let n_{x_i} be the number of customers choosing table x_i. Then the utility function of customer i is given by $U(T_{x_i}, n_{x_i})$, where $U(\cdot)$ is a decreasing function of n_{x_i} for the effect of negative network externality. Supposing customers sequentially arrive and request for seats from these K tables, social learning can be included. Under such a situation, how the customers learn the true size of tables and then choose the right tables to enhance their own dining experience? The above table selection problem is an example of Chinese restaurant game, which draws immediate parallels with decentralized channel selection in cognitive radio networks. For details of Chinese restaurant game, we refer readers to [17] and [18], where the relation between Chinese restaurant game and Chinese restaurant process is also described.

1.1. Related works

There is an extensive research on the topic of opportunistic spectrum access [2–8]. We refer readers to [19] for an overview. In recent years, a few works have been done in the framework of multi-armed bandit (MAB), where an SU's decision on whether to sense new channels in the hope of obtaining better availability (i.e., exploration) or to transmit over the current channel (i.e., exploitation) is investigated. While it is assumed in the early works [6,20] that the channel availability statistics are known and there is no competition among SUs, current researches [21,22] abandon such assumptions when studing multiple SUs' optimal channel selection strategy. Considering that channel states are selected by adversary (thus non-stochastic), the authors of [23] proposed joint channel sensing, probing, and accessing schemes and proved that the proposed schemes can achieve almost optimal throughput. In all these schemes SUs learn independently, reducing the overhead of information exchange but at the same time missing the chance of social learning which may shorten the time to converge and enhance learning accuracy. Moreover, taking network externality into consideration, it is more natural and necessary to formulate the decision making process of rational and selfish SUs into a game.

Two closely-related strategic game models which incorporate the negative externality of spectrum resource are global game and congestion game, in which the payoff of each player depends on the resource it chooses and the number of players choosing the same resource. In [24], considering that competitive optimal behavior of the secondary system is a function of the prior probability distribution of spectrum hole occupancy, channel quality and observation noise, the decentralized dynamic spectrum access problem among SUs was discussed in the framework of interacting multivariate global game. In [25], multiple heterogenous SUs' competition for transmission on idle primary channels was finally modeled as a singleton congestion game. Since all players in these games make decisions simultaneously, there is no social learning involved.

Some researchers have made great efforts to introduce learning and signaling into global game and congestion game. Angeletos et al. [26,27], for example, examined how learning influences the dynamics of coordination in a global game of regime change, where a status quo is abandoned once a sufficiently large fraction of agents attacks it. However, the network externality considered in these works are positive. In [28], the authors investigated learning through multiplicative updates in congestion games, while in [29] the authors proposed a stochastic learning automata (SLA) based channel selection algorithm. Nevertheless, due to simultaneous decision, players in these games still can only learn from their private information or action-reward history. Moreover, the multiplicity of equilibria of these games is analyzed only in simplified models where players are homogenous. A general study incorporating both effects of social learning and negative network externality is still limited, especially in the case where heterogeneous players make decisions sequentially.

Recently, Wang et al. [17,18] studied a sequential decision scenario where a certain number of intelligent users make wise decisions by taking advantages of other users' experiences through learning to avoid competitions from huge crowds. By introducing the strategic behavior into the non-strategic Chinese restaurant process, the authors proposed a new game called Chinese restaurant game. Jiang et al. [30] then extended the game model to the dynamic scenario where users may arrive at or leave the network at any time, and directly apply the proposed dynamic Chinese restaurant game to cognitive radio networks to demonstrate the effectiveness and efficiency of the game from simulations. Our work differs from these works in two aspects. First, instead of building a general framework and using cognitive radio network as background for simulations, we apply the model to specific cognitive radio scenario where we perform extensive theoretical analysis and get meaningful insights into the impact of important factors, such as initial belief, sensing accuracy and channel quality, on SUs' access decisions. Second, while [17,18] and [30] only consider the homogeneous case where users have the same valuation about the same resource, our work takes into account the heterogeneous characteristic of users, i.e., the quantity of the same channel is variant for different SUs.

In our prior work [1], we formulated SUs' decision making problem in opportunistic spectrum access as a Chinese restaurant game, where belief threshold is not considered and we only derived SUs' optimal access actions and the corresponding action regions under different initial beliefs and sensing accuracies, and studied the property of SUs' expected action for the two-user two-channel scenario. While in this paper, we modify the system model to a more practical one by adding a belief threshold for SU's access channel selection, where we get totally different results on SUs' optimal access actions and the corresponding action regions. Based on the modified model, we further investigate and obtain more insight on the impact of channel quality on SUs' actions. We then extend our discussion on the two-user two-channel scenario to the multi-user multi-channel scenario. What's more, we propose a Backward Induction Algorithm to find the best strategy for each SU, and prove theoretically that the algorithm will lead to sub-game perfect Nash equilibrium (NE) of the proposed game and under certain conditions the NE is unique. The convergence and computational complexity of the algorithm is analyzed as well.

1.2. Contributions

The novelty and technical contributions of this work are summarized as follows.

- Different from previous works which focused on simultaneous spectrum access decision, we address the sequential access decision problem, which is more realistic but more difficult, for cognitive radio networks. The decision making process in such a problem is modeled as a Chinese restaurant game. In our proposed game, all SUs first simultaneously sense the channels to estimate the channel state and then decide sequentially which channel to access. Once an SU has selected a channel, it reveals its access decision as well as the sensing result through a common control channel which can be overheard by all other SUs. Since an SU can exploit the information collected by sensing and learning, it is able to build a global view of the channel state of the system and choose the optimal channel for access, not necessarily constrained to the one it has sensed.
- We propose a backward induction algorithm to find the optimal action for each SU in the system. Furthermore, we prove that the algorithm will result in a sub-game NE of the proposed game and under certain conditions the NE is unique.
- Taking the heterogeneous characteristic of SUs into consideration, we further get meaningful insights into the interactions among SUs. Specifically, we derive SUs' optimal access actions and study the impact of initial belief, sensing accuracy and channel quality on their decisions. We also discuss the SUs' expected actions before knowing the sensing results and study some of the general properties of expected actions under different channel qualities. Furthermore, we extend the discussion for two-user two-channel scenario to multi-user multi-channel scenario. Note that the theoretical analysis is non-trivial. Due to the highly non-linear characteristic of expected utility function defined in the game, no closed-form solution for the scenario under consideration can be obtained. The heterogeneous characteristic of SUs makes it more difficult to analyze SUs' actions mathematically since the results can be significantly different when SUs' preferences are differently given. Simulations are finally performed to verify

Table 1
Notations.

a_m	The access action of SU_m
a_m^*	The optimal access action of SU_m
$\mathbf{a}^* = \{a_1^*, \ldots, a_M^*\}$	M SUs' optimal action profile
$b_{m,k}$	The belief, i.e., the probability, of channel H_k being vacant from the perspective of SU_m
$\mathbf{b} = (b_{0,1}, \ldots, b_{0,K})$	The prior belief with $b_{0,k}$ being the prior belief on channel H_k
c_m	The sensing action of SU_m
$g_{m,k}$	SU_m's average channel gain in H_k
$\mathbf{G} = (g_{m,k} \mid \forall m, k)$	The channel quality of the system
H_k	The Kth licensed channel
θ_k	The state of channel H_k
$\boldsymbol{\theta}$	Channel state of the system
Θ	Space of channel state
$f(s\mid\theta_k)$	Predefined distribution of result s when sensing channel H_k
p	Sensing accuracy
$s_{m,k} \in \{s^+, s^-\}$	The signal obtained via SU_m's sensing channel H_k
$R_{m,k}(\mathbf{G})$	The maximal rate SU_m can obtain by accessing channel H_k
$R_{m,k}(\mathbf{G}, n)$	The rate that SU_m can obtain when it shares channel H_k with $n-1$ other SUs
$\mathbf{h}_m$	The set of signals revealed by SUs before SU_m
$\mathbf{n}_m$	Vector used to record the number of SUs choosing each channel before SU_m
$v_{m,k}$	The number of SUs choosing H_k after SU_m, including SU_m itself
$U_{m,k}(\cdot)$	SU_m's utility of accessing channel H_k
$\bar{U}_{m,k}(\cdot)$	The expected utility of SU_m accessing channel H_k
$BE_m(\cdot)$	SU_m's best response
$\Phi_i, i = 0, 1, 2$	Action regions for SU_1 in the two-channel two-user case
$\Psi_i, i = 0, 1, 2, 3$	Action regions for SU_2 in the two-channel two-user case
$\varphi_m(\cdot)$	The expected action of SU_m

the theoretical results and evaluate the performance of the proposed scheme in terms of social welfare.

Note that our model is quite general and its application is not restricted to the channel selection problem but can also be deployed in other problems in cognitive radio networks with sequential decision making and negative network externality, such as relay selection with limit transmission power and/or backhaul capability. Beyond that, the proposed model can be applied to many other fields, such as service selection in cloud computing, deal selection on Groupon website in online social networking and WiFi access point selection in wireless networking.

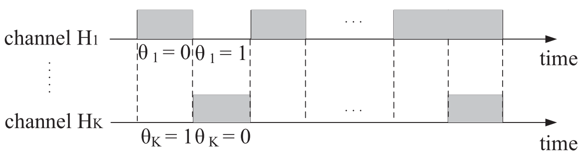
In the rest of the paper, Section 2 describes in details our system model for the cognitive radio networks and formulate the decision making problem as a Chinese restaurant game. The existence of Nash equilibrium is proved and the algorithm to find the Nash equilibrium is also proposed in this section. In Section 3, we analyze the impact of prior belief, sensing accuracy and the channel quality on SUs' decision in the two-user two-channel scenario while in Section 4, we extend the analysis to the multi-user multi-channel scenario. Finally, we present the simulation results in Section 5 and draw conclusions in Section 6. A list of the key mathematical symbols used in this paper is shown in Table 1.

2. Chinese restaurant game model of cognitive radio system

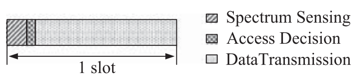
2.1. System model

In this paper, we consider a primary system with K licensed channels, H_k, $k \in \mathcal{K} = \{1, 2, \ldots, K\}$, as shown in Fig. 1. All the channels in the system are synchronized and slotted, and each is owned by one PU. Within each slot, the channel state of the system is defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)$ with state space Θ. θ_k, the state of channel H_k, can be 0 or 1 according to the PU's activity. Here, "0" stands for the channel being occupied by the PU while "1" means that the channel is vacant. The PUs' access pattern $\Pr(\theta_k)$, $\forall k \in \mathcal{K}$, are independent with each other and considered as prior probabilities.

Suppose that there are M SUs, i.e., SU_m, $m \in \mathcal{M} = \{1, 2, \ldots, M\}$, searching vacant channels for transmission. Since SUs are not licensed users, they can only access the channels when the PUs are not present. In such a case, SUs need to perform sensing before accessing the channels. All SUs can independently perform sensing using energy detection [2]. The sensing result, which is a binary signal $s \in \{s^+, s^-\}$, represents the observation or estimation of the state of the sensed channel. The positive signal s^+ indicates that the channel is vacant while the negative signal s^- stands for



(a) Synchronized and slotted channels



(b) The slot structure

Fig. 1. The system model.

the channel being occupied by the PU. Considering the uncertainty in sensing process due to physical propagation effect, e.g. fading or shadowing, we further assume that s follows a predefined Binomial distribution, i.e., $f(s = \theta_k|\theta_k) = p$ and $f(s \neq \theta_k|\theta_k) = 1 - p$, $\forall k \in \mathcal{K}$. Here $p \in (0, 1)$ can be regarded as the system parameter and is referred to sensing accuracy in the rest of this paper.

As shown in Fig. 1(b), in our model, one slot is further divided into three sub-slots. In the first sub-slot, M SUs perform sensing. Although our model can be extended to multiple spectrum sensing, in this paper, we assume that an SU only senses one channel. After sensing, each SU can select one vacant channel for transmission due to the hardware limitation [31]. Therefore, in the second sub-slot, SUs sequentially make the access decisions based on the information they have collected,[1] and report their decisions as well as sensing results within a predefined time through a dedicated common control channel which can be overheard by all other SUs.[2] Note that the sequential decision order will cause the unbalance information availability among SUs and affect their decisions and utilities. For the sake of fairness, we assume that the decision order is randomized and is different at each time slot. In the third sub-slot, SUs transmit their data through the channels they selected. If more than one SU choose the same channel, they can share the channel through but not limited to Carrier Sense Multiple Access (CSMA), Time Division Multiple Access (TDMA) or Code Division Multiple Access (CDMA).

Note that since SUs are uncertain about their future decision orders, they only care about the expected utility at the current slot. Moreover, an SU's utility is obtained within the third sub-slot which will be normalized to 1 in the following analysis. In such a case, how the three sub-slots are divided will not affect the SUs' decisions. However, it should be pointed out that the optimal sub-slot allocation is an important issue for certain scenarios and previous works focusing on this topic can be referred in [33] and [34].

2.2. Utility function

Let $g_{m,k}(t)$ be SU$_m$'s average channel gain in H_k at slot t, which generally varies over the slots but remains invariant within each slot [38,39]. Since in this paper we mainly discuss SUs' decisions in a specific time slot, the time index t is thereby dropped. With $g_{m,k}$, we define $\mathbf{G} = (g_{m,k}|m \in \mathcal{M}, k \in \mathcal{K})$ as the channel quality of the system in a slot, which can be achieved by channel estimation [35,40] and known to every SU.[3] Then given $\mathbf{G}$, the maximal rate SU$_m$ can obtain by accessing channel H_k is $R_{m,k}(\mathbf{G}) = \log_2(1 + \frac{g_{m,k}P_m}{N_0^2})$, where P_m is SU$_m$'s transmission power and N_0^2 is the variance of additive white Gaussain noise. We assume that all SUs use the same power to transmit and all the noises have equal variance. Moreover, let $R_{m,k}(\mathbf{G}, n)$ be the rate that SU$_m$ can obtain when it shares channel H_k with $n-1$ other SUs. Note that $R_{m,k}(\mathbf{G}, n)$ should be a decreasing function in terms of n due to interference or less access opportunity, and its exact form is determined by how these SUs share the channel.

Definition 1 (Preferential channel)**.** Channel H_k is the preferential channel of SU$_m$ if $H_k = \arg \max_{H_k \in \{H_1, \ldots, H_K\}} R_{m,k}(\mathbf{G})$.

We use the transmission throughput as SUs' utilities. Assuming the length of the third sub-slot is normalized to 1, the utility of SU$_m$ accessing channel H_k can be defined as

$$U_{m,k}(\mathbf{G}, \theta_k, N_k) = R_{m,k}(\mathbf{G}, N_k)\mathbf{1}(\theta_k = 1), \quad (1)$$

where $\mathbf{1}(\cdot)$ is the indicator function and N_k is the final number of SUs that choose to access channel H_k.

From (1) we can see that the SU's utility is determined by the channel quality, the channel state and the number of SUs who share this channel. However, due to sensing error and sequential decision, the true channel state θ_k and the value of N_k, $\forall k \in \mathcal{K}$, are unknown to SU$_m$ in the decision making phrase (the second sub-slot). Therefore, instead of maximizing utility defined in (1), an SU learns the true channel state and estimates the number of users who will eventually share the channel with it, aiming at maximizing the expected utility defined in the following section. Such a decision making process involving learning and prediction is thereby formulated as a Chinese restaurant game [17].

2.3. Chinese restaurant game

Let $c_m \in \mathcal{C}$ and $a_m \in \mathcal{A}$ be the sensing and access actions of SU$_m$, where $\mathcal{C} = \{1, 2, \ldots, K\}$ and $\mathcal{A} = \{0, 1, 2, \ldots, K\}$ are the sensing and access action sets that SUs may choose from. Here, "0" in $\mathcal{A}$ is used to represent that an SU will choose none of the K channels.

We use the concept of belief to describe an SU's estimation on the channel state. Specifically, let belief $b_{m,k}$ be the probability that channel H_k is vacant from the perspective of SU$_m$. Let $\mathbf{b} = (b_{0,1}, b_{0,2}, \ldots, b_{0,K})$ with $b_{0,k}$ being the prior belief on channel H_k which reflects the kth PU's historical access pattern. For simplicity, we assume that all SUs have a common prior belief on the same channel. Let $s_{m,k} \in \{s^+, s^-\}$ be the signal obtained via SU$_m$'s sensing channel H_k while $\bar{s}_{m,k} \in \{s^+, s^-\} \backslash s_{m,k}$ be the complement signal of $s_{m,k}$. Assume that all the signals are independent when conditioning on the channel state. With the collected signals $\mathcal{T}_k = \{s_{i,k}|c_i = k, 1 \leq i \leq m\}$, SU$_m$ can update its belief on H_k with Bayesian rule as follows:

[1] Note that the time an SU needs to make decision can be very short, if the channel gain is modeled as a Finite-State Markov chain [36], the computation of optimal strategies is conducted off-line and the results are put into a look-up table, to which an SU refers for a quick decision.

[2] The report can be sent in a message with format < i, s, j >, where $s \in \{0, 1\}$ is the sensing result, and i and j, both represented in $\log K$ bits, are the channels that the SU has sensed and will access, respectively. As a result, $2\log K + 1$ bits are needed for one message and the total overhead is $M(2\log K + 1)$ bits, if modulation is not taken into account. Such an overhead is acceptable even in a real secondary system where M and K is on the order of tens [31,32].

[3] We assume that the channel quality information is broadcasted by SUs via a wireless control channel every coherence interval. Then the time for information propagation is less than 0.002 ms assuming the average distance between two SUs is 5 km and the total delay for channel quality information exchange is less than 0.002M ms, while the report overhead will be nKM bits if each report needs n bits.

$$b_{m,k}(b_{0,k}, p, \mathcal{T}_k) = \frac{\prod_{s\in\mathcal{T}_k} f(s|\theta_k = 1)b_{0,k}}{\prod_{s\in\mathcal{T}_k} f(s|\theta_k = 1)b_{0,k} + \prod_{s\in\mathcal{T}_k} f(s|\theta_k = 0)(1 - b_{0,k})}. \quad (2)$$

Suppose there is a belief threshold $\underline{b}_k$ for each channel H_k, $\forall k \in \mathcal{K}$. If an SU's belief on the channel is below $\underline{b}_k$, it is not allowed to access H_k. Assume that the belief thresholds for all channels are the same, i.e., $\underline{b}_k = \underline{b}$, $\forall k \in \mathcal{K}$. Then $\mathcal{K}_m = \{k|b_{m,k} \geq \underline{b}, \forall k \in \mathcal{K}\}$ is denoted as the set of channels that SU_m is allowed to access.[4]

Besides learning the channel state, an SU also needs to predict the decisions of the subsequent SUs due to the existence of negative network externality. Recall that an SU will report its sensing result and access decision after it makes the access decision. Let $\mathbf{h}_m = (s_{1,c_1}, s_{2,c_2}, \ldots, s_{m-1,c_{m-1}})$ be the signals revealed by the SUs before SU_m and $\mathbf{n}_m = (n_{m,1}, n_{m,2}, \ldots, n_{m,K})$ with $n_{m,k}$ being the number of SUs choosing channel H_k before SU_m. If we denote $v_{m,k}$ be the number of SUs choosing H_k after SU_m, including SU_m itself, when all the information $\mathbf{b}$, p, $\mathbf{G}$, $\mathbf{n}_m$, $\mathbf{h}_m$ and s_{m,c_m} are available, the expected utility of SU_m accessing channel H_k over different channel state $\boldsymbol{\theta}$ and all the possible number of $v_{m,k}$, is defined as

$$\bar{U}_{m,k}(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m}) = \begin{cases} \sum_{\boldsymbol{\theta}\in\Theta} \sum_{x=1}^{M-m+1} [Pr(\boldsymbol{\theta}|\mathbf{b}, p, \mathbf{h}_m, s_{m,c_m}) \\ \cdot Pr(v_{m,k} = x|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m}, a_m = k, \boldsymbol{\theta}) \cdot \\ U_{m,k}(\mathbf{G}, \theta_k, n_{m,k} + x)], \quad \text{if } k \in \mathcal{K}_m, \\ 0, \quad \text{otherwise}, \end{cases} \quad (3)$$

where $Pr(\boldsymbol{\theta}|\cdot)$ is the probability that the system state is $\boldsymbol{\theta}$ conditioning on $\mathbf{b}$, p, $\mathbf{h}_m$ and s_{m,c_m}, and $Pr(v_{m,k} = x|\cdot)$ is the conditional probability of $v_{m,k}$ being x. Therefore, given $\mathbf{b}$, p, $\mathbf{G}$, $\mathbf{n}_m$, $\mathbf{h}_m$ and s_{m,c_m}, SU_m's best response, i.e., the optimal access action that maximizes the expected utility, can be defined as

$$BE_m(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m}) = \begin{cases} \arg\max_{k\in\mathcal{K}_m} \bar{U}_{m,k}(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m}), & \text{if } \mathcal{K}_m \neq \varnothing \\ 0, & \text{otherwise}. \end{cases} \quad (4)$$

2.4. Backward induction algorithm

In this section, we study how to find the best response for each SU_m, $m \in \mathcal{M}$.

From (3) we can see that to compute the expected utility of accessing certain channel, say H_k, SU_m needs to estimate the channel state, i.e., $Pr(\boldsymbol{\theta}|\mathbf{b}, p, \mathbf{h}_m, s_{m,c_m})$. Given all the condition, $Pr(\boldsymbol{\theta}|\cdot)$ can be computed by following the Bayesian rule as

$$Pr(\boldsymbol{\theta}|\mathbf{b}, p, \mathbf{h}_m, s_{m,c_m}) = \frac{Pr(\boldsymbol{\theta})Pr(\mathbf{h}_m, s_{m,c_m}|\boldsymbol{\theta})}{\sum_{\theta'\in\Theta} Pr(\theta')Pr(\mathbf{h}_m, s_{m,c_m}|\theta')}, \quad (5)$$

where $Pr(\boldsymbol{\theta}) = \prod_{k\in\mathcal{K}} Pr(\theta_k)$ due to independent state of each channel, and $Pr(\mathbf{h}_m, s_{m,c_m}|\boldsymbol{\theta})$ depends on how the signals are observed, i.e., $f(s|\theta_k)$, $\forall k \in \mathcal{K}$.

SU_m also needs to guess how many SUs after it will make the same decision, i.e., $v_{m,k}$, and the corresponding probability $Pr(v_{m,k}|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m}, a_m, \boldsymbol{\theta})$. To reach this target, the backward induction technology is used as follows.

Notice from (4) that there may exist more than one signal that leads SU_{m+1} to choose certain channel when $\mathbf{b}$, p, $\mathbf{G}$, $\mathbf{n}_{m+1}$ and $\mathbf{h}_{m+1}$ are given. Then let $\mathcal{S}_{m+1,k}$, $1 \leq m < M$, $k \in \mathcal{K}^+ = \mathcal{K} \cup \{0\}$, be the signal space with signal from which SU_{m+1} will choose channel H_k, we have

$$\mathcal{S}_{m+1,k}(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_{m+1}, \mathbf{h}_{m+1}) = \{s_{m+1,c_{m+1}}|BE_{m+1}(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_{m+1}, \mathbf{h}_{m+1}, s_{m+1,c_{m+1}}) = k, \forall c_{m+1} \in \mathcal{C}\}. \quad (6)$$

In other words, given the necessary conditions mentioned before, SU_{m+1} will choose channel H_k if and only if its sensing result $s_{m+1,c_{m+1}} \in \mathcal{S}_{m+1,k}$. In such a case, SU_{m+1}'s access action can be predicted according to the signal distribution $f(s)$ by

$$Pr(a_{m+1} = k|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_{m+1}, \mathbf{h}_{m+1}) = \sum_{s\in\mathcal{S}_{m+1,k}} f(s). \quad (7)$$

Moreover, according to the definition of $v_{m,k}$, we have

$$v_{m,k} = \begin{cases} 1 + v_{m+1,k}, & \text{if } a_m = k, \\ v_{m+1,k}, & \text{otherwise}. \end{cases} \quad (8)$$

Note that for the last SU, i.e., SU_M, $Pr(v_{M,k} = x|\cdot)$ can be easily derived as

$$Pr(v_{M,k} = 1|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_M, \mathbf{h}_M, s_{M,c_M}, a_M, \boldsymbol{\theta}) = \begin{cases} 1, & \text{if } a_M = k, \\ 0, & \text{otherwise}. \end{cases} \quad (9)$$

Then the recursive form of $Pr(v_{m,k} = x|\cdot)$ can be expressed as

$$Pr(v_{m,k} = x|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m}, a_m, \boldsymbol{\theta}) = \begin{cases} \sum_{u\in\mathcal{K}^+} \sum_{s\in\mathcal{S}_{m+1,u}} Pr(v_{m+1,k} = x-1|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_{m+1}, \mathbf{h}_{m+1}, s_{m+1,c_{m+1}} \\ \quad = s, a_{m+1} = u, \boldsymbol{\theta})f(s|\boldsymbol{\theta}), \ a_m = k, \\ \sum_{u\in\mathcal{K}^+} \sum_{s\in\mathcal{S}_{m+1,u}} Pr(v_{m+1,k} = x|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_{m+1}, \mathbf{h}_{m+1}, s_{m+1,c_{m+1}} \\ \quad = s, a_{m+1} = u, \boldsymbol{\theta})f(s|\boldsymbol{\theta}), \ a_m \neq k, \end{cases} \quad (10)$$

where $\mathbf{h}_{m+1} = (\mathbf{h}_m, s_{m,c_m})$, and $\mathbf{n}_{m+1} = \mathbf{n}_m + \mathbf{e}_{a_m}$ with $\mathbf{e}_{a_m}$ being a standard basis vector whose a_mth coordinate is 1 while other coordinates are 0 (if $a_m > 0$) or $\mathbf{n}_{m+1} = \mathbf{n}_m$ (if $a_m = 0$).

With (3), (4), (5) and (10), the best response of SU_m can be calculated. The proposed backward induction algorithm is summarized in Algorithm 1.

[4] From Figs. 5–7 in Section 5, we can see that both the collision probability with PUs and the spectrum opportunity for SUs reduce with the belief threshold. Therefore, there is a tradeoff between the collision probability and the spectrum opportunity with different thresholds. One possible way to select the belief threshold is to maximize the spectrum opportunity with the constraint that the collision probability should be smaller than a certain threshold.

Algorithm 1 Find the optimal access decision a_m^* for SU_m by backward induction.

Given $\mathbf{b}$, p, $\mathbf{G}$, $\mathbf{n}_m$, $\mathbf{h}_m$ and s_{m,c_m}.
If SU $m = M$, then

- calculate $Pr(v_{m,k}|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m}, \boldsymbol{\theta})$, $\forall k \in \mathcal{K}$, with (9)
- compute $Pr(\boldsymbol{\theta}|\mathbf{b}, p, \mathbf{h}_m, s_{m,c_m})$, $\forall k \in \mathcal{K}$, using (5)
- $\mathcal{K}_m = \{k|b_{m,k}(\mathbf{b}, p, \mathbf{h}_m, s_{m,c_m}) \ge \underline{b}, \forall k \in \mathcal{K}\}$
- calculate $\bar{U}_{m,k}(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m})$, $\forall k \in \mathcal{K}_m$, with (3)
- $BE_m(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m})$ is obtained by (4)

else // $1 \le m < M$

- get $BE_M(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_M, \mathbf{h}_M, s_{M,c_M})$ by following the previous steps
- for i=M-1:m
  - $\mathbf{h}_{i+1} = (\mathbf{h}_i, s_{i,c_i})$
  - calculate $Pr(\boldsymbol{\theta}|\mathbf{b}, p, \mathbf{h}_i, s_{i,c_i})$ using (5)
  - $\forall k \in \mathcal{K}^+$
    - $\mathbf{n}_{i+1} = \begin{cases} \mathbf{n}_i + \mathbf{e}_k, k \neq 0 \\ \mathbf{n}_i, k = 0 \end{cases}$
    - calculate $\mathcal{S}_{i+1,k}(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_{i+1}, \mathbf{h}_{i+1})$ with (6)
    - calculate $Pr(v_{i,k}|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_i, \mathbf{h}_i, s_{i,c_i}, \boldsymbol{\theta})$ with (10)
  - $\mathcal{K}_i = \{k|b_{i,k}(\mathbf{b}, p, \mathbf{h}_i, s_{i,c_i}) \ge \underline{b}, \forall k \in \mathcal{K}\}$
  - calculate $\bar{U}_{i,k}(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_i, \mathbf{h}_i, s_{i,c_i})$, $\forall k \in \mathcal{K}_i$, using (3)
  - obtain $BE_i(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_i, \mathbf{h}_i, s_{i,c_i})$ by (4)
- end for

End If
$a_m^* \leftarrow BE_m(\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m})$
End

From Algorithm 1 we can see that the algorithm guarantees to converge since there is finite SUs and finite channels in the system, and the number of recursion of the algorithm is $\mathcal{O}(M)$, which is quite efficient even when the size of secondary system is large. Within each recursion, the maximum computation of an SU is $(2K^2)^{M+1}$, Therefore, the total computational complexity of the algorithm is $\mathcal{O}((2K^2)^{M+1})$, i.e., exponential in both the space size of channel state and signal and linear in the number of SUs. However, by introducing belief threshold, the complexity has been reduced by nearly $(100 * \underline{b})$%. In addition, considering the channel gain can be modeled as a Finite-State Markov chain [36], the computation can be conducted off-line and the results can be put into a look-up table, to which an SU refers for decision in no time.

In the next section, we will show that Algorithm 1 leads to sub-game perfect NE [37] for the proposed game and under certain conditions the NE is unique.

2.5. Subgame perfect Nash equilibrium

In this section, we first give the formal definitions of NE, subgame and subgame perfect NE.

Definition 2 (Nash equilibrium)**.** Given $(\mathbf{b}, p, \mathbf{G})$, M SUs' action profile $\mathbf{a}^* = \{a_1^*, a_2^*, \ldots, a_M^*\}$ is a Nash equilibrium if and only if $\forall m \in \mathcal{M}$, $a'_m \in \mathcal{A} \setminus a_m^*$, a_m^* satisfies

$$\bar{U}_{m,a_m^*} \ge \bar{U}_{m,a'_m}, \tag{11}$$

where $\bar{U}_{m,a_m^*}$ and $\bar{U}_{m,a'_m}$ are defined in (3).

Definition 3 (Subgame)**.** A subgame of the M SUs' channel access game consists of the following three elements: (1) it starts from the ith SU, $i = 1, 2, \ldots, M$; (2) it has the initial condition $\mathbf{b}$, p, and $\mathbf{G}$; (3)$\forall i \le m \le M$, there are current observations $\mathbf{n}_m$, $\mathbf{h}_m$ and s_{m,c_m} for SU_m.

Definition 4 (Subgame Perfect NE)**.** An NE is a subgame perfect NE if and only if it is a NE for every subgame [37].

With the above definitions, we show in Theorem 1 that the action profile derived by Algorithm 1 is a subgame perfect NE.

Theorem 1. *The action profile* $\mathbf{a}^* = \{a_1^*, a_2^*, \ldots, a_M^*\}$ *which is derived by Algorithm 1 is a subgame perfect NE.*

Proof. We first show that a_m^* is the best response of the mth SU in the subgame which starts from the ith SU, $1 \le i \le M$.

If $m = M$, we can have from Algorithm 1 that $a_M^* = k$ if for any $k' \in \mathcal{A} \setminus k$, there is $\bar{U}_{M,k} \ge \bar{U}_{M,k'}$. This means a_M^* is the best response when all other SUs' strategies are given.

If $m < M$, suppose $a_m^* = k_1$ is derived by $BE_m(\cdot)$ and SU_m's other choices are $a'_m = k'_1$, $k'_1 \neq k_1$. Then for any $k'_1 \in \mathcal{A} \setminus k_1$, we have $\bar{U}_{m,k_1} \ge \bar{U}_{m,k'_1}$, which means that SU_m has no incentive to deviate from a_m^* given the prediction of other SUs' decision. Therefore, $a_m^* = BE_m(\cdot)$ is SU_m's best response in the subgame starting with the ith SU. In summary, since the statement is true for any $1 \le i \le M$, $\mathbf{a}^* = \{a_1^*, a_2^*, \ldots, a_M^*\}$ is NE for the subgame starting with the ith SU. According to the Definition 4, we can finally conclude that Theorem 1 is true. □

Note that the subgame perfect NE may not be unique. For example, if the equality of (11) holds for certain SUs, then each of such SUs has at least two channels that can bring it maximal expected utility. Choosing any one of such channels will lead to a different subgame perfect NE. However, the NE can be unique as stated in Theorem 2.

Theorem 2. *If (11) strictly holds for all SUs, then NE* $\mathbf{a}^* = \{a_1^*, a_2^*, \ldots, a_M^*\}$ *is unique.*

Proof. We prove Theorem 2 by contradiction. Suppose that when (11) strictly holds for all SUs, there exist two Nash equilibria $\mathbf{a}^* = \{a_1^*, a_2^*, \ldots, a_M^*\}$ and $\mathbf{a}^{*'} = \{a_1^{*'}, a_2^{*'}, \ldots, a_M^{*'}\}$ where $a_m^* \neq a_m^{*'}$ for some $m \in \mathcal{M}$ since $\mathbf{a}^*$ and $\mathbf{a}^{*'}$ are different Nash equilibria. In such a case, we have $\bar{U}_{m,a_m^*} > \bar{U}_{m,a_m^{*'}}$ and $\bar{U}_{m,a_m^{*'}} > \bar{U}_{m,a_m^*}$. Obviously, the above two inequalities cannot hold at the same time. Therefore, the NE is unique when (11) strictly holds for all SUs. □

3. Two-user two-channel scenario

In this section, we analyze the interactions among SUs for the two-user two-channel scenario, i.e., $K = 2$ and $M = 2$. We first derive SUs' optimal access actions under different $\mathbf{b}$ and

p by assuming the channel quality $\mathbf{G}$ and the sensing results are given. Then, we discuss the SUs' expected actions before knowing the sensing results. Finally, we study some general properties of SUs' expected actions under different channel qualities.

3.1. Optimal actions and action regions with sensing results

To give more insight of the proposed approach, we assume that SUs' prior belief on both channels are the same, i.e., $b_{0,1} = b_{0,2} = b_0$. Note that such an assumption is reasonable since PUs in these two channels can have similar access patterns. We also give the definition of action region as follows.

Definition 5 (Action region)**.** An action region is an area in the plane of b_0 and p, where an SU adopts a specific strategy.

Then as discussed in the previous section, we use backward induction to derive SUs' optimal actions. In the following, we first analyze SU$_2$'s strategies and obtain the corresponding optimal access action regions given the sensing results as described in Theorem 3.

Theorem 3. *Suppose H_i is the preferential channel of SU_2. When $c_1 \neq c_2$ and $s_{1,c_1} \neq s_{2,c_2}$, or $c_1 = c_2$ and $s_{1,c_1} = s_{2,c_2}$, there are four possible action regions , $\Psi_i, i = 0, 1, 2, 3$, for SU_2 on the plane of b_0 and p as follows.*

- $\Psi_0 = \{(b_0, p)|b_{2,i} < \underline{b}, b_{2,-i} < \underline{b}\}$ *with the optimal action* $a_2^* = 0$,
- $\Psi_1 = \bar{\Psi}_0^i \cup (\{(b_0, p)|\frac{b_{2,i}}{b_{2,-i}} > \frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G})}\} \cap \Phi_0 \cap \bar{\Psi}_0^{i,-i}) \cup (\{(b_0, p)|\frac{b_{2,i}}{b_{2,-i}} > \frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G},2)}\} \cap \bar{\Phi}_0 \cap \bar{\Psi}_0^{i,-i})$ *with the optimal action* $a_2^* = i$,
- $\Psi_2 = \bar{\Psi}_0^{-i} \cup (\{(b_0, p)|\frac{b_{2,-i}}{b_{2,i}} > \frac{R_{2,i}(\mathbf{G})}{R_{2,-i}(\mathbf{G})}\} \cap \Phi_0 \cap \bar{\Psi}_0^{i,-i}) \cup (\{(b_0, p)|\frac{b_{2,-i}}{b_{2,i}} > \frac{R_{2,i}(\mathbf{G})}{R_{2,-i}(\mathbf{G},2)}\} \cap \bar{\Phi}_0 \cap \bar{\Psi}_0^{i,-i})$ *with the optimal action* $a_2^* = -i$,
- $\Psi_3 = \Omega - \Psi_0 - \Psi_1 - \Psi_2$ *with the optimal action* $a_2^* = -a_1$,

where $-i \in \mathcal{K}\backslash i$, $\Phi_0 = \{(b_0, p)|b_{1,i} < \underline{b}, b_{1,-i} < \underline{b}\}$, $\bar{\Phi}_0 = \Omega - \Phi_0$, $\bar{\Psi}_0^i = \{(b_0, p)|b_{2,i} \geq \underline{b} > b_{2,-i}\}$, $\bar{\Psi}_0^{-i} = \{(b_0, p)|b_{2,-i} \geq \underline{b} > b_{2,i}\}$, $\bar{\Psi}_0^{i,-i} = \{(b_0, p)|b_{2,i} \geq \underline{b}, b_{2,-i} \geq \underline{b}\}$, $\Omega = \{(b_0, p)|1 > b_0 > 0, 1 > q > 0\}$, *and* $b_{1,i}$, $b_{1,-i}$, $b_{2,i}$ *and* $b_{2,-i}$ *are given by (2).*

On the other hand, when $c_1 = c_2$ and $s_{1,c_1} \neq s_{2,c_2}$, or $c_1 \neq c_2$ and $s_{1,c_1} = s_{2,c_2}$, there will be only two possible optimal actions on the whole plane of b_0 and p, that is, Ψ_0 and $\bar{\Psi}_0 = \Omega - \Psi_0$.

Proof. See proof in Appendix A. □

Based on SU$_2$'s optimal action regions, we can analyze SU$_1$'s strategies and derive the corresponding optimal action regions as follows.

Theorem 4. *Suppose H_j is the preferential channel of SU_1. Then, SU_1's optimal actions and the corresponding action regions, $\Phi_i, i = 0, 1, 2$, can be written as follows.*

- Φ_0 *with the optimal action* $a_1^* = 0$,
- $\Phi_1 = \cup_d \phi_d$ *with the optimal action* $a_1^* = j$,
- $\Phi_2 = \cup_d \bar{\phi}_d$ *with the optimal action* $a_1^* = -j$,

where $-j \in \mathcal{K}\backslash j$, $d \in \mathcal{D} = \{0, 1, 2, 3\}$, ϕ_d *and* $\bar{\phi}_d$ *are respectively defined as*

$$\phi_d = \left(\left\{(b_0, p)\middle|\frac{\bar{b}_{1,j}(b_0, p, \mathbf{G}, s_{1,c_1}|\Psi_d)}{\bar{b}_{1,-j}(b_0, p, \mathbf{G}, s_{1,c_1}|\Psi_d)} > \frac{R_{1,-j}(\mathbf{G})}{R_{1,j}(\mathbf{G})}\right\} \cap \bar{\Phi}_0^{j,-j}\right) \cup \bar{\Phi}_0^j, \tag{12}$$

and

$$\bar{\phi}_d = \left(\left\{(b_0, p)\middle|\frac{\bar{b}_{1,j}(b_0, p, \mathbf{G}, s_{1,c_1}|\Psi_d)}{\bar{b}_{1,-j}(b_0, p, \mathbf{G}, s_{1,c_1}|\Psi_d)} < \frac{R_{1,-j}(\mathbf{G})}{R_{1,j}(\mathbf{G})}\right\} \cap \bar{\Phi}_0^{j,-j}\right) \cup \bar{\Phi}_0^{-j}, \tag{13}$$

where $-j \in \mathcal{K}\backslash j$, $\bar{\Phi}_0^j = \{(b_0, p)|b_{1,j} \geq \underline{b} > b_{1,-j}\}$, $\bar{\Phi}_0^{-j} = \{(b_0, p)|b_{1,-j} \geq \underline{b} > b_{1,j}\}$, *and* $\bar{\Phi}_0^{j,-j} = \{(b_0, p)|b_{1,j} \geq \underline{b}, b_{1,-j} \geq \underline{b}\}$

Proof. See proof in Appendix B. □

From the analysis of SUs' optimal strategies and the corresponding action regions in Theorems 3 and 4, we have the following observations.

- When SUs have the same preferential channel, they will share the preferential channel in region ϕ_1 and share the non-preferential channel in region $\bar{\phi}_3$.
- When SUs have their own preferential channel, respectively, they will share SU$_1$'s preferential channel in region ϕ_3 and share SU$_2$'s preferential channel in region $\bar{\phi}_1$.
- Given s_{1,c_1}, SU$_1$'s action will be independent from the actual signal SU$_2$ receives.

3.2. Expected actions without sensing results

In the previous section, we derive SUs' optimal strategies and the corresponding action regions given the sensing results. In this section, we will analyze the symmetric property of SUs' expected actions without the sensing results. Note that the expected action can be served as the SUs' prior information about their optimal actions before actually performing sensing.

Definition 6 (Expected action)**.** For any $(b_x, p_y) \in \{(b_0, p)\}$, the expected action of SU$_i$, $i \in \{1, 2\}$, is defined as

$$\varphi_i(b_x, p_y) = \sum_{s \in \{\mathbf{h}_i, s_{i,c_i}\}} Pr(s|b_x, p_y) \cdot a_i(s, b_x, p_y), \tag{14}$$

where s is the signal(s) SU$_i$ collected, $Pr(s|b_x, p_y)$ is the probability of receiving s under b_x and p_y, and $a_i(s, b_x, p_y)$ is SU$_i$'s action when it receives s.

Let s_1 and s_2 denote the sensing signals SU$_1$ and SU$_2$ get respectively. To show the property of the expected actions, we first characterize, in Lemmas 1 and 2, the property of SUs' optimal actions and action regions when receiving opposite sensing results.

Lemma 1. *Given $\mathbf{G}$, SU_2 will choose the same optimal strategy in the action region $\Phi_d(b_0, p)$ with sensing results (s_1, s_2) and the action region $\Phi_d(b_0, 1-p)$ with sensing results $(\bar{s}_1, \bar{s}_2)$.*

Proof. See proof in Appendix C. □

Lemma 2. *Given* $\mathbf{G}$, *SU_1 will choose the same optimal strategy in the action region $\phi_d(b_0, p)$ with sensing results s_1 and the action region $\phi_d(b_0, 1-p)$ with sensing results $\bar{s}_1$.*

Proof. See proof in Appendix D. □

With Lemmas above, we are ready to show the symmetric property of *SUs*' expected actions.

Theorem 5. *Given* $\mathbf{G}$, *the expected actions of SU_2 are symmetrical to $p = 0.5$.*

Proof. From (14), $\forall(b_x, p_y) \in \{(b_0, p)\}$, SU_2's expected action is defined as

$$\begin{aligned}\varphi_2(b_x, p_y) = &\ Pr(s_{1,c_1}, s_{2,c_2}|b_x, p_y) \cdot a_2(s_{1,c_1}, s_{2,c_2}, b_x, p_y)\\ &+ Pr(\bar{s}_{1,c_1}, \bar{s}_{2,c_2}|b_x, p_y) \cdot a_2(\bar{s}_{1,c_1}, \bar{s}_{2,c_2}, b_x, p_y)\\ &+ Pr(s_{1,c_1}, \bar{s}_{2,c_2}|b_x, p_y) \cdot a_2(s_{1,c_1}, \bar{s}_{2,c_2}, b_x, p_y)\\ &+ Pr(\bar{s}_{1,c_1}, s_{2,c_2}|b_x, p_y) \cdot a_2(\bar{s}_{1,c_1}, s_{2,c_2}, b_x, p_y).\end{aligned} \tag{15}$$

From Lemma 1, we know that if there is a region Ψ_d, $d \in \mathcal{D}$, where SU_2 has specific strategy when it receives $(s_1, s_2) \in \{(s_{1,c_1}, s_{2,c_2}), (\bar{s}_{1,c_1}, s_{2,c_2}), (s_{1,c_1}, \bar{s}_{2,c_2}), (\bar{s}_{1,c_1}, \bar{s}_{2,c_2})\}$, then the region for SU_2 having the same strategy when receiving $(\bar{s}_1, \bar{s}_2)$ is Ψ'_d. Moreover, with Lemma 1, we have $\Psi'_d(b_0, p) = \Psi_d(b_0, 1-p)$. In such a case, $\forall(b_x, p_y) \in \Psi_d$, there is $(b_x, 1-p_y) \in \Psi'_d$, thus we have

$$a_2(s_1, s_2, b_x, p_y) = a_2(\bar{s}_1, \bar{s}_2, b_x, 1-p_y). \tag{16}$$

Furthermore, since $Pr(s_1, s_2|b_0, p)$ represents the probability of receiving (s_1, s_2) given b_0 and p, we have

$$Pr(s_1, s_2|b_x, p_y) = Pr(\bar{s}_1, \bar{s}_2|b_x, 1-p_y). \tag{17}$$

Then by substituting (16) and (17) into (15), we have

$$\varphi_2(b_x, p_y) = \varphi_2(b_x, 1-p_y), \tag{18}$$

which means the expected actions of SU_2 is symmetrical to $p = 0.5$. □

Theorem 6. *Given* $\mathbf{G}$, *the expected actions of SU_1 are symmetrical to $p = 0.5$.*

Proof. From (14), $\forall(b_x, p_y) \in \{(b_0, p)\}$, SU_1's expected action is defined as

$$\begin{aligned}\varphi_1(b_x, p_y) = &\ Pr(s_{1,c_1}|b_x, p_y) \cdot a_1(s_{1,c_1}, b_x, p_y)\\ &+ Pr(\bar{s}_{1,c_1}|b_x, p_y) \cdot a_1(\bar{s}_{1,c_1} b_x, p_y).\end{aligned} \tag{19}$$

From Lemma 2, we know that if there is a region ϕ_d, $d \in \mathcal{D}$, where SU_1 has specific strategy when receiving $s_1 \in \{s_{1,c_1}, \bar{s}_{1,c_1}\}$, there is a corresponding region ϕ'_d for SU_1 having the same strategy when receiving $\bar{s}_1$. Moreover, with Lemma 2, we have $\phi'_d(b_0, p) = \phi_d(b_0, 1-p)$. Then, $\forall(b_x, p_y) \in \phi_d$, there is $(b_x, 1-p_y) \in \phi'_d$, and thus we have

$$a_1(s_1, b_x, p_y) = a_1(\bar{s}_1, b_x, 1-p_y). \tag{20}$$

Note that for $\bar{\phi}_d$, $d \in \mathcal{D}$, we also have the above conclusions.

Furthermore, since $Pr(s_1|b_0, p)$ represents the probability of receiving s_1 given b_0 and p, we have

$$Pr(s_1|b_x, p_y) = Pr(\bar{s}_1|b_x, 1-p_y). \tag{21}$$

Then by substituting (20) and (21) into (19), we have

$$\varphi_1(b_x, p_y) = \varphi_1(b_x, 1-p_y), \tag{22}$$

which means the expected actions of SU_1 are symmetrical to $p = 0.5$. □

The symmetry property of SU's expected action is mainly due to the fact that an original positive signal obtained with a low sensing accuracy p can be treated as a negative one obtained with a high sensing accuracy $1-p$, and vice versa. With such a property, an SU is able to obtain its expected actions on the whole b_0 and p plane by calculating the expected actions on any half of the plane with $p < 0.5$ or $p > 0.5$, and thus accelerate its process of decision making.

3.3. Expected actions under different channel quality

In this section, we discuss the impact of the channel quality on SUs' expected actions. Since in regions Ψ_0 and Φ_0, SUs' believes do not depend on the channel quality but the signals they received, here, we constrain our discussion on the region $\bar{\Phi}_0 \cap \bar{\Psi}_0$, where both SUs are allowed to access at least one of the channels. We first show in Theorem 7 that if $\mathbf{G}$ satisfies some conditions, then SU_1's expected action only depends on its sensing action regardless SU_2's sensing action. In all the following, the superscript (i, j) is used to denote the case $c_1 = i$ and $c_2 = j$.

Theorem 7. $\forall(b_x, p_y) \in \bar{\Phi}_0 \cap \bar{\Psi}_0$, *we have* $\varphi_1^{(i,i)}(b_x, p_y) = \varphi_1^{(i,-i)}(b_x, p_y)$, *if* $\mathbf{G} \in \mathcal{G} = \{\mathbf{G}|\chi_1(\mathbf{G}) \cap \chi_2(\mathbf{G}) \cap \chi_3(\mathbf{G})\}$, *where χ_1, χ_2 and χ_3 are functions of* $\mathbf{G}$ *defined in (s29), (s30) and (s31) of [41], respectively.*

Proof. See proof in the Supplementary [41]. □

In Theorem 7, $\chi_1(\mathbf{G})$ represents the SUs' preferential channel and $\chi_2(\mathbf{G})$ represents the difference between the gain of preferential channel and non-preferential channel, which determines SU_1's access action regions $\chi_3(\mathbf{G})$ when b_0 and p are given. From Theorem 7, we can see that if the channel quality $\mathbf{G}$ is one of the element in the set $\mathcal{G}$, SU_1's access action regions will be the same as long as it senses the same channel and gets the same sensing result. In such a case, SU_1's expected actions are the same in two scenarios where SU_1 senses a specific channel while SU_2 senses the same channel or a different channel, i.e., SU_1's expected action is independent of SU_2's sensing action.

Next, we will show in Theorem 8 that when $\mathbf{G}$ satisfies some conditions, SU_2's expected action will be the same once it senses a different channel from SU_1 regardless which exact channel it performs sensing.

Theorem 8. $\forall(b_x, p_y) \in \bar{\Phi}_0 \cap \bar{\Psi}_0$, *we have* $\varphi_2^{(i,j)}(b_x, p_y) = \varphi_2^{(j,i)}(b_x, p_y)$, $i \neq j$, *if* $\mathbf{G} \in \mathcal{G} = \{\mathbf{G}|\chi_1(\mathbf{G}) \cap \chi_2(\mathbf{G}) \cap \chi_4(\mathbf{G})\}$, *where χ_1, χ_2 and χ_4 are functions of* $\mathbf{G}$ *defined in (s29), (s30) and (s41) of [41], respectively.*

Proof. See proof in the Supplementary [41]. □

Similarly, in Theorem 8, SU_2's access action regions $\chi_4(\mathbf{G})$ is determined by the SUs' preferential channel $\chi_1(\mathbf{G})$ and the difference between the gain of preferential channel and non-preferential channel $\chi_2(\mathbf{G})$. From Theorem 8, we can see that if the channel quality $\mathbf{G}$ is one of the element in the set $\mathcal{G}$, SU_2's access action regions will be the same for the two scenarios, where SU_2 senses a different channel from SU_1 and collects the same information for the channels. Note that SU_1 may have different action regions in the two aforementioned scenarios. However, with $\mathbf{G} \in \mathcal{G}$, the changes happen in the regions Ψ_1 and/or Ψ_3, where SU_2's access action will not change according to SU_1's action. In such a case, SU_2's expected action are the same in the two scenarios for any given b_0 and p.

4. Multi-user multi-channel scenario

In this section, we extend our discussion for two-user two-channel scenario to multi-user multi-channel scenario,[5] i.e., $K > 2$ and $M > 2$, where we derive the optimal action for each user and the corresponding action regions with sensing results.

From previous analysis, we know that for an SU, for example SU_m, based on the quality of all channels $\mathbf{G}$, the sensing results revealed by the previous SUs $\mathbf{n}_m$, the number of SUs that have accessed the channels before it $\mathbf{h}_m$, and the sensing result obtained by itself s_{m,c_m}, the SU can estimate the state of all channels, predict the actions of SUs after it, and then make its decision of accessing the channel. For simplicity, let $\Gamma_m = \{\mathbf{n}_m\} \times \{\mathbf{h}_m\} \times \{s_{m,c_m}\}$ be the set of possible information that can be collected by SU_m, and $|\Gamma_m|$ be the total number of elements in Γ_m. We also denote $\phi_{m,k}(\mathbf{G}, \boldsymbol{\gamma}_{m,\tau})$ be the action region for SU_m accessing H_k when given $\mathbf{G}$ and $\boldsymbol{\gamma}_{m,\tau} \in \Gamma_m, \forall \tau \in \{1, 2, \ldots, |\Gamma_M|\}$.

In our model, the number of the total SUs is assumed to be finite, and therefore we still can adopt backward induction and first derive the optimal action and action region for the last SU.

4.1. Optimal action and action region for SU_M

Since SU_M is the last SU, it does not need to predict but just makes the access decision based on the collected information, i.e., $\boldsymbol{\gamma}_{M,\tau} \in \Gamma_M = \{\mathbf{n}_M\} \times \{\mathbf{h}_M\} \times \{s_{M,c_M}\}, \ \forall \tau \in \{1, 2, \ldots, |\Gamma_M|\}$. In such a case, given $\mathbf{G}$ and $\boldsymbol{\gamma}_{M,\tau}$, SU_M's best response can be derived according to (4) as

$$a_M^* = BE_M(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{M,\tau}) = \begin{cases} \arg\max\limits_{k \in \mathcal{K}_M} \bar{b}_{M,k}(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{M,\tau}) R_{M,k}, & \text{if } \mathcal{K}_M \neq \varnothing \\ 0, & \text{otherwise}, \end{cases} \tag{23}$$

where $\bar{b}_{M,k}(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{M,\tau}) = b_{M,k}(\mathbf{b}, p, \boldsymbol{\gamma}_{M,\tau}) \alpha_{M,k}(\mathbf{G}, n_{M,k} + 1)$ and $b_{M,k}$ is given by (2).

Therefore, the action region for the optimal action $a_M^* = i$ can be computed by

$$\phi_{M,i}(\mathbf{G}, \boldsymbol{\gamma}_{M,\tau}) = \left\{(\mathbf{b}, p) | \bar{b}_{M,i} R_{M,i} = \max_{k \in \mathcal{K}^+} \bar{b}_{M,k}(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{M,\tau}) R_{M,k}\right\}. \tag{24}$$

[5] Here, multi channels refers that there are more than 2 channels in the primary system among which, however, each SU is allowed to choose only one to sense.

4.2. Optimal action and action region for SU_m with $m < M$

In this section, we derive the optimal access action and action region for SU_m $(m < M)$ by establishing the recurrence relation between SU_m and SU_{m+1}.

Note that the information SU_m collects is $\boldsymbol{\gamma}_{m,\tau} \in \Gamma_m$, $\tau \in \{1, 2, \ldots, |\Gamma_m|\}$. To find out the signal space $\mathcal{S}_{m+1,k}$ and further estimate SU_{m+1}'s access action, SU_m needs to consider the access action it may take, i.e., a_m, and the signal SU_{m+1} may obtain, i.e., $s_{m+1,c_{m+1}}$. However, based on $\boldsymbol{\gamma}_{m,\tau}$, for any $a_m \in \mathcal{A}$ and $s_{m+1,c_{m+1}} \in \{s_+, s_-\}$, the best response of SU_{m+1} also depends on which action region the given $\mathbf{b}$ and p are in. If we denote $\Lambda_{m+1} = \{\bigcap_{\tau,k_\tau} \phi_{m+1,k_\tau}(\mathbf{G}, \boldsymbol{\gamma}_{m+1,\tau}) | \tau = 1, 2, \ldots, |\Gamma_{m+1}|, k_\tau \in \mathcal{K}\}$ be the set of SU_{m+1}'s action region where it has specific optimal action. Then $\forall \lambda_{m+1} \in \Lambda_{m+1}$ and for any given $\boldsymbol{\gamma}_{m,\tau}$, a_m and $s_{m+1,c_{m+1}}$, SU_{m+1}'s best response can be obtained as

$$BE_{m+1}(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{m+1,\tau} | \lambda_{m+1}) = BE_{m+1}(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{m,\tau}, a_m, s_{m+1,c_{m+1}} | \lambda_{m+1}). \tag{25}$$

With (25), we have $\mathcal{S}_{m+1,k}$ according to (6) and then the recursive form of $Pr(v_{m,k} = x | (\mathbf{b}, p) \in \lambda_{m+1}, \mathbf{G}, \boldsymbol{\gamma}_{m,\tau}, a_m, \boldsymbol{\theta})$ according to (10).

Finally, SU_m's best response can be derived according to (4) as

$$BE_m(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{m,\tau} | \lambda_{m+1}) = \begin{cases} \arg\max\limits_{k \in \mathcal{K}} \bar{b}_{m,k}(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{m,\tau} | \lambda_{m+1}) R_{m,k}, & \text{if } \mathcal{K}_m \neq \varnothing \\ 0, & \text{otherwise}, \end{cases} \tag{26}$$

where

$$\bar{b}_{m,k}(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{m,\tau} | \lambda_{m+1}) = \sum_{\boldsymbol{\theta} \in \Theta} \sum_{x=0}^{M-m+1} [\alpha_{m,k}(\mathbf{G}, n_{m,k} + x) Pr(\boldsymbol{\theta} | \mathbf{b}, p, \boldsymbol{\gamma}_{m,\tau}) \cdot Pr(v_{m,k} = x | (\mathbf{b}, p) \in \lambda_{m+1}, \mathbf{G}, \boldsymbol{\gamma}_{m,\tau}, a_m = k, \boldsymbol{\theta})]. \tag{27}$$

Let $\phi_{m,i}(\mathbf{a}_s, \mathbf{G}, \boldsymbol{\gamma}_{m,\tau} | \lambda_{m+1})$ be the action region where SU_m will access H_i conditioning on region λ_{m+1}, we have

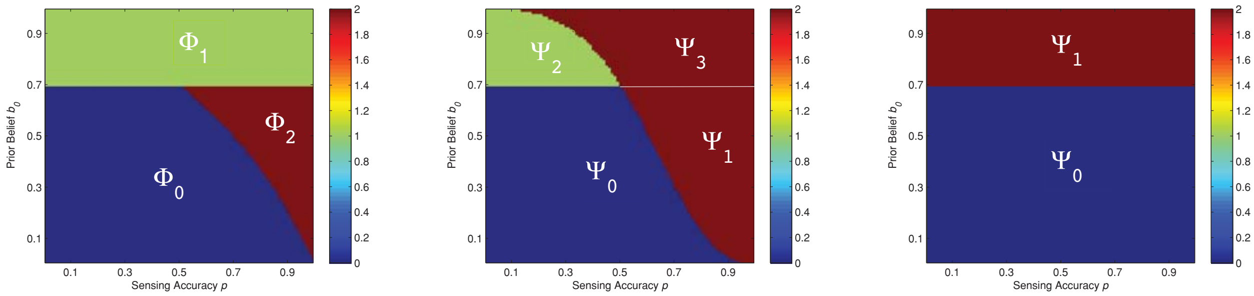
$$\phi_{m,i}(\mathbf{G}, \boldsymbol{\gamma}_{m,\tau} | \lambda_{m+1}) = \left\{(\mathbf{b}, p) | \bar{b}_{m,i} R_{m,i} = \max_{k \in \mathcal{K}_m} \bar{b}_{m,k}(\mathbf{b}, p, \mathbf{G}, \boldsymbol{\gamma}_{m,\tau} | \lambda_{m+1}) R_{m,k}\right\}. \tag{28}$$

Then, given $\mathbf{G}$ and $\boldsymbol{\gamma}_{m,\tau} \in \Gamma_m$, the action region for the optimal action $a_m^* = i$ can be computed by

$$\phi_{m,i}(\mathbf{G}, \boldsymbol{\gamma}_{m,\tau}) = \bigcup_{\lambda_{m+1} \in \Lambda_{m+1}} \phi_{m,i}(\mathbf{G}, \boldsymbol{\gamma}_{m,\tau} | \lambda_{m+1}). \tag{29}$$

5. Simulation results

In this section, we first verify the optimal action, action region and expected action of SUs in the proposed game theoretic framework for two-channel two-user scenario. Then we verify the NE of the proposed approach. Finally, we evaluate the system performance for multi-user multi-channel scenario in terms of social welfare.



(a) $SU_1, s_1 = s^+$ (b) $SU_2, s_1 = s^+, s_2 = s^+$ (c) $SU_2, s_1 = s^+, s_2 = s^-$

Fig. 2. Action regions of SU_1 and SU_2 with $c_1 = 2$, $c_2 = 2$, $\mathbf{G} = [10, 1; 1, 10]$ and $b_{01} = b_{02} = b_0$.

In the following, we assume that all SUs have the same initial belief b_0 and sensing accuracy p, and b_0 and p are respectively set to one of the values in {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}. The maximum transmission power of SU_m is set to $P_m = 1$, $m \in \mathcal{M}$, the noise variance of each channel is set to $N_0^2 = 1$, and the belief threshold is set to $\underline{b} = 0.7$. Since the simulation results are similar for different MAC schemes, here we only show those with CSMA where the probability of successful transmission decreases as the number of SUs selecting the same channel increases. In such a case, the SU's utility is defined as [25]

$$U_{m,k}^{CSMA}(\mathbf{G}, \theta_k, N_k) = R_{m,k}(\mathbf{G})\mathbf{1}(\theta_k = 1)/N_k. \tag{30}$$

5.1. Optimal actions and action region with sensing results

We consider a network composed by 2 PUs and 2 SUs, i.e., $K = 2$, $M = 2$. Moreover, SUs' transmitters and receivers are deployed in such places that $g_{11} = g_{22} = 10$ and $g_{12} = g_{21} = 1$, that is, channel H_1 is SU_1's preferential channel while channel H_2 is SU_2's preferential channel. Fig. 2 shows the optimal action regions when both SUs sense channel H_2, i.e., $c_1 = 2$ and $c_2 = 2$. Since both SUs sense channel H_2, SUs' believes on channel H_1 remain unchanged while their believes on channel H_2 will be updated according to the sensing results.

From Fig. 2(b) and (c), we can see that there are four action regions when both sensing results are positive, while there are only two action regions when one of the sensing results is positive and the other is negative. Such phenomena verify the theoretical results in Theorem 3. As shown in Fig. 2(b), in Ψ_0, SU_2's initial believes on both channels are below the belief threshold. With two positive signals for H_2, SU_2's belief on H_2 is pulled down to a lower level in region $p < 0.5$ where the positive signals work as negative ones. In the $p > 0.5$ part of Ψ_0, even with positive signals, the initial belief is too low to be improved to the belief threshold. As a result, SU_2 is not allowed to access either of the two channels. In Ψ_1, SU_2's initial belief is not that low, so the belief on H_2 can be improved to or even higher than the belief threshold by two positive signals in $p > 0.5$, with which SU_2 has a stronger belief that H_2 is empty when compared to H_1. Moreover, accessing H_2 can bring a larger payoff due to the higher channel gain, even when considering SU_1 may also access the same channel. Therefore, SU_2 chooses H_2. While in Ψ_2, SU_2's initial believes on two channels are above the belief threshold. However, SU_2's belief on H_2 is pulled down to below the belief threshold due to the fact the positive signals are obtained under the condition of $p < 0.5$. In such a case, SU_2 has no choice but access the non-preferential channel H_1, sharing H_1 with SU_1. Nevertheless, when (b_0, p) shifts from region Ψ_2 to Ψ_3, the increase of belief on H_2 enables SU_2 to choose from both channels. Although SU_2's belief on H_1 is still stronger than that on H_2, the gain from higher belief can no longer compensate the loss from low channel gain and sharing the channel with SU_1. Consequently, the best strategy for SU_2 is to access a different channel from SU_1.

The action regions of SU_1 are shown in Fig. 2(a). We can see that there are three possible action regions for SU_1, which verifies the results in Theorem 4. Recalling that H_1 is SU_1's preferential channel, in action region Φ_1 where SU_1's initial believes on H_1 is over the belief threshold, SU_1 chooses to access H_1 because the higher channel gain can bring a larger payoff, even when the belief on H_2 may be higher.

5.2. Expected actions without sensing results

In this section, we evaluate the SUs' expected actions for the case $K = 2$, $M = 2$, and analyze the impact of channel quality on SUs' expected actions by studying the same case as described in the previous section. The outcomes of the simulation are shown in Fig. 3.

From Fig. 3 we can see that the expected actions of both SU_1 and SU_2 are symmetrical to $p = 0.5$, which verifies Theorems 5 and 6. We can also see that SU_1's expected actions in Fig. 3(a) and (c) are the same as those in Fig. 3(b) and (d) respectively. Such a phenomenon is because the channel gain satisfies the condition in Theorem 7, due to which SU_1's expected action only depends on its sensing action regardless SU_2's sensing action. Furthermore, since the channel gain also meets the requirement in Theorem 8, SU_2's expected action will be the same once it senses a different channel from SU_1 regardless which exact channel it performs sensing. Such a conclusion is verified by Fig. 3(b) and (c) where SU_2's expected actions are the same.

5.3. Nash equilibrium

In this section, we verify that the proposed approach leads to Nash equilibrium, i.e., any deviation to other strategies will lead to a utility loss. To this aim, we compare our approach, i.e., CRG, with three existing strategies, i.e., CUCB [42], MYOPIC [25], and CRG_WO [1].

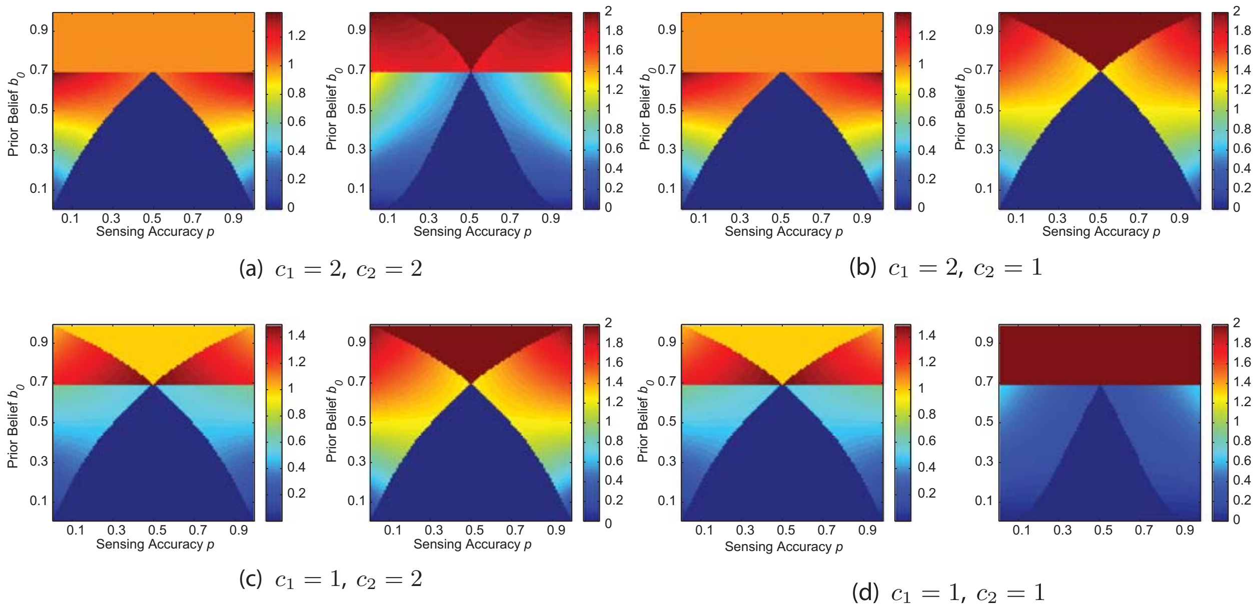


Fig. 3. Expected actions of SU_1 and SU_2 with $\mathbf{G} = [10,1;1,10]$.

In the CUCB scheme, an SU, say SU_m, learns the availability of channel H_j, which is reflected by the Upper Confidence Bound index $B_{m,j}$, through its own reward information and reward knowledge shared by other SUs [42]. For fair comparison, we assume all SUs broadcast their reward information. With the spectrum availability information, SU_m choose to sense channel H_j with probability given as follow and access the channel when the channel is found free:

$$\pi^m(j) = \frac{e^{B_{m,j}}}{\sum_{m=0}^{M} e^{B_{m,j}}}. \tag{31}$$

In the MYOPIC strategy where the belief threshold is added for fair comparison, a myopic SU makes the decision according to its own signal and the current number of SUs choosing the same channel [31]. Therefore, the objective of the SU under MYOPIC strategy is to choose the access channel that can maximize its current expected utility given by

$$a_{am}^{MYOPIC} = \begin{cases} \arg\max\limits_{k\in\mathcal{K}_m} b_{m,k} U_{m,k}(\mathbf{G}, \theta_k = 1, n_{m,k} + 1), & \text{if } \mathcal{K}_m \neq \varnothing \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

In the CRG_WO strategy which is very similar to the CRG strategy except that there is no belief threshold for SU's access channel selection, an SU makes the access decision to maximize the expected utility [1], i.e.,

$$\begin{aligned} a_{am}^{CRGWO} = \max_{k\in\mathcal{K}} \sum_{\boldsymbol{\theta}\in\Theta} \sum_{x=1}^{M-m+1} & [Pr(\boldsymbol{\theta}|\mathbf{b}, p, \mathbf{h}_m, s_{m,c_m}) \\ & \cdot Pr(v_{m,k} = x|\mathbf{b}, p, \mathbf{G}, \mathbf{n}_m, \mathbf{h}_m, s_{m,c_m}, a_m = k, \boldsymbol{\theta}) \\ & \cdot U_{m,k}(\mathbf{G}, \theta_k, n_{m,k} + x)]. \end{aligned} \tag{33}$$

Assume that among the SUs, SU_2 may adopt one of the following four strategies: CRG, CUCB, MYOPIC and CRG_WO. The rest of SUs all use the proposed strategy CRG. We measure the ratio between the utility generated by any three other strategies and the utility generated by CRG, and the results of scenario where all the channels are homogeneous for SUs are shown in Fig. 4. From Fig. 4, we can see that the ratio is smaller than or equal to 1 for any b_0 and p in the region where SU_2 is allowed to access at least one of the channels, which means that the proposed strategy is indeed a Nash equilibrium.

5.4. System performance

In the following, we study the system performance of secondary system in terms of social welfare, i.e., the sum of all SUs' utilities, in the system with multiple SUs and multiple channels. Besides the schemes mentioned in the previous section, a centralized strategy named "CEN" is considered here as a benchmark for comparison and evaluation. In the centralized strategy, sensing signals from all the SUs is collected and the best access decision which can maximize total social welfare is made as follows:

$$a_a^{CEN} = \begin{cases} \arg\max\limits_{a_a\in A_a} b(s_1, \ldots, s_m) U(\mathbf{G}, \boldsymbol{\theta}, a_a), & \text{if } \mathcal{K} \neq \varnothing, \\ 0, & \text{otherwise,} \end{cases} \tag{34}$$

where a_a is the access schedule for all the SUs while A_a is the set of all the possible schedules, $b(\cdot)$ is the belief for system state based on all the signals collected, and $\mathcal{K}$ is the set of channels to whom the belief is above the belief threshold.

We also verify the performance of the proposed scheme under different MAC protocols, including CSMA, TDMA and CDMA. For TDMA where the time of transmission decreases as the number of SUs selecting the same channel increases and CDMA where the interference increases as the number of SUs selecting the same channel increases, the SU's utilities

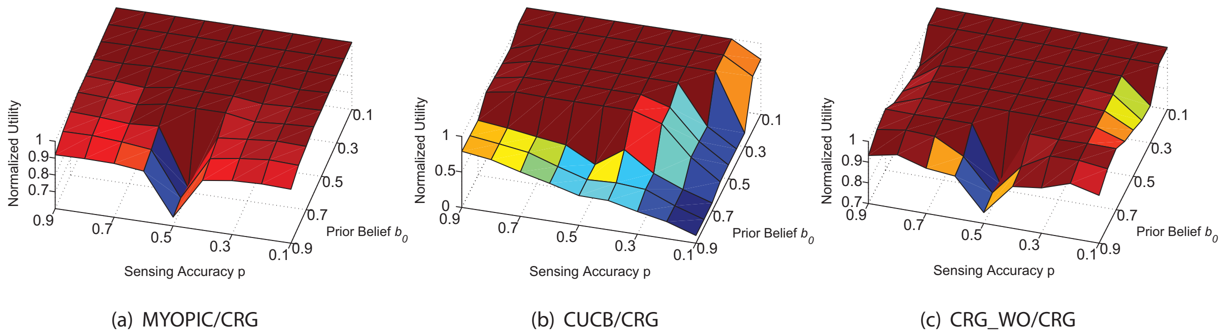


Fig. 4. Normalized utility of SU_2 with $\mathbf{G} = [10{,}10;10{,}10;10{,}10;10{,}10;10{,}10]$.

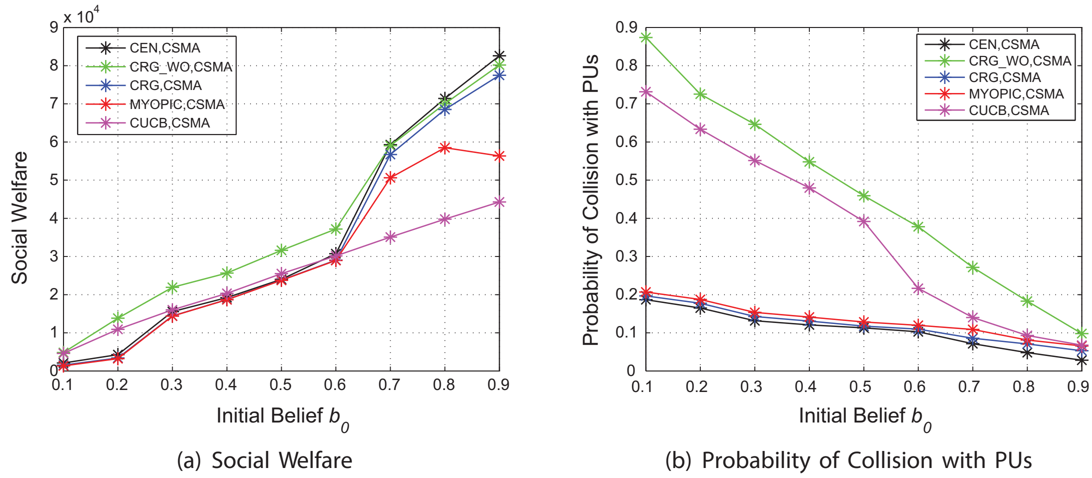


Fig. 5. Performance of schemes with CSMA and $\mathbf{G} = [7{,}10{,}1;5{,}10{,}4;10{,}5{,}1;7{,}10{,}1;5{,}10{,}4;10{,}5{,}1;7{,}10{,}1;5{,}10{,}4;10{,}5{,}1;7{,}10{,}1;5{,}10{,}4;10{,}5{,}1]$ when $p = 0.9$.

are respectively defined as

$$U_{m,k}^{TDMA}(\mathbf{G}, \theta_k, N_k) = R_{m,k}(\mathbf{G})\mathbf{1}(\theta_k = 1)/N_k, \tag{35}$$

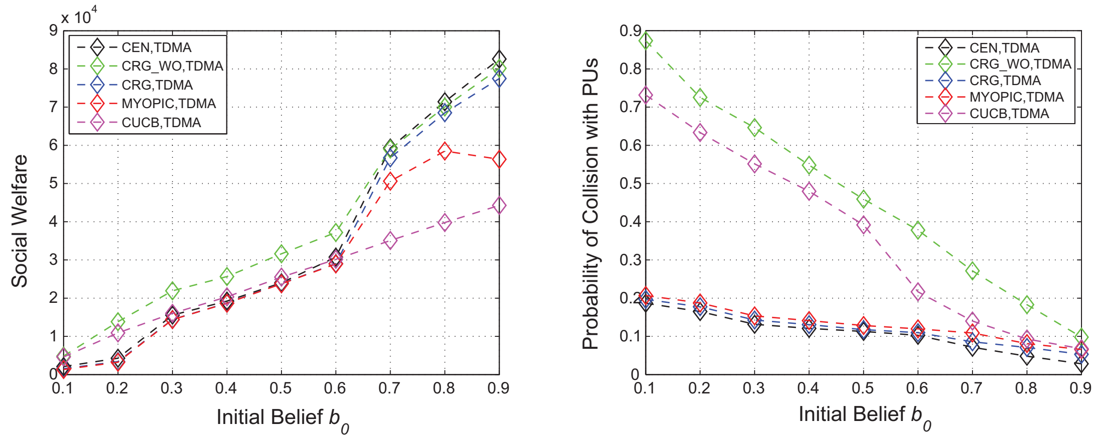
and

$$\begin{aligned} U_{m,k}^{CDMA}(\mathbf{G}, \theta_k, N_k) &= R_{m,k}(\mathbf{G}, N_k)\mathbf{1}(\theta_k = 1) \\ &= \log_2\left(1 + \frac{g_{m,k}P_m}{(N_k - 1) * N_0^2}\right)\mathbf{1}(\theta_k = 1). \end{aligned} \tag{36}$$

In the simulation, all SUs in the system will adopt the same strategy with the same MAC protocol. For CSMA, TDMA and CDMA, the performance of a cognitive radio network with 12 SUs and 3 channels are shown in Figs. 5,6 and 7, respectively. From Fig. 5 to Fig. 7, we can observe that performances of our model with different MAC protocols are very similar. In such a case, in the following we only focus on analyzing the results with one of the MAC protocols, say CSMA in Fig. 5.

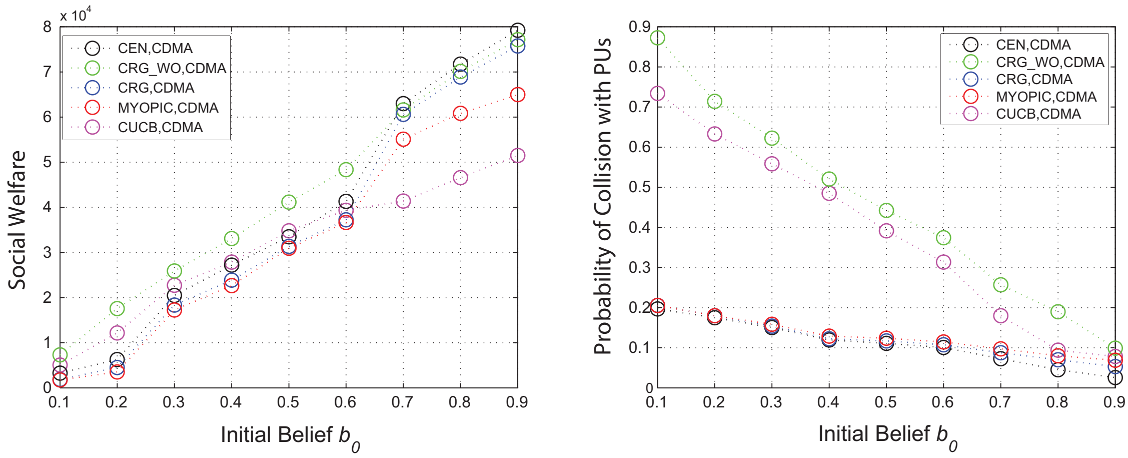
From Fig. 5(a), we can see that on an average the social welfare with CRG has been increased to 13.47% and 29.56% compared to that with MYOPIC and CUCB, respectively. That's because in CUCB, although the added reward information is used to improve the sensing decision making behavior, SUs make their access decisions purely based on their own sensing result which could be easily affected by the sensing accuracy, and thus have the lowest social welfare. By also considering actions of previous SUs, the system performance with MYOPIC scheme is improved due to the conflict among SUs being partly avoided. In our proposed scheme, through learning sensing signals as well as access actions revealed by previous SUs and estimating the actions of subsequent SUs, an SU successfully avoids the conflict with the PU and other SUs and therefore outperforms SUs in MYOPIC and CUCB. Moreover, from Fig. 5, we can see that the improvement of the social welfare with CRG is more significant in area where b_0 is large, which indicates that the proposed approach is efficient for practical system where the probability of channel vacancy is high.

Notice that the social welfare of CRG is less than that of CRG_WO for all b_0 and that of CUCB when $b_0 < 0.7$. This is because with belief threshold, SUs with CRG in these cases give up the access chance to avoid collision with PUs. As we can observe in Fig. 5(b), the ratio of collision between SUs and PUs with CRG in the aforementioned cases is much less when compared to that with CRG_WO and CUCB, while it is almost the same as that with MYOPIC due to the same belief threshold.



(a) Social Welfare (b) Probability of Collision with PUs

Fig. 6. Performance of schemes with TDMA and $\mathbf{G} = [7,10,1;5,10,4;10,5,1;7,10,1;5,10,4;10,5,1;7,10,1;5,10,4;10,5,1;7,10,1;5,10,4;10,5,1]$ when $p = 0.9$.



(a) Social Welfare (b) Probability of Collision with PUs

Fig. 7. Performance of schemes with CDMA and $\mathbf{G} = [7,10,1;5,10,4;10,5,1;7,10,1;5,10,4;10,5,1;7,10,1;5,10,4;10,5,1;7,10,1;5,10,4;10,5,1]$ when $p = 0.9$.

Finally, our proposed CRG works almost as well as the centralized scheme CEN, since the performance gap between them is less than 10%, as illustrated in Fig. 5.

6. Conclusion

In this paper, we successfully address the distributed access channel selection problem for SUs by formulating SUs' decision making process problem in opportunistic spectrum access as a Chinese restaurant game. With the proposed game theoretic approach, SUs can make better decisions and achieve better performance through learning from others and estimating others' decisions. We theoretically derive SUs' optimal access actions and the corresponding action regions under different initial conditions. We also study some general properties such as symmetric property of SUs' expected actions under different channel qualities for the two-user two-channel scenario. Furthermore, we extend the discussion for two-user two-channel scenario to multi-user multi-channel scenario. Simulation results verify our theoretic results and demonstrate the effectiveness and efficiency of the proposed scheme.

Based on the achievements in this paper, it is more interesting to study scenario where SUs can sense/select more than one channel. Furthermore, since the sensing strategy will influence SUs' access behaviors as well as the system overall performance, how to choose the proper channel(s) to sense is one of the important problems for future work, where an algorithm with low overhead and computational complexity is very necessary for finding the optimal solutions. We are certainly interested in further pursuing these interesting research directions.

Acknowledgments

This paper was supported by National Natural Science Foundation of China (Nos. 61501041, 61002011 and 61372116), National Research Foundation for the Doctoral Program of Higher Education of China (No. 20120005110007), China 973 Program (No. 2012CB315801), Beijing Natural Science Foundation (No. 4122034) and Open Foundation of State Key Laboratory (No. ISN16-08).

Appendix A. Proof of Theorem 3

Theorem 3. *Suppose H_i is the preferential channel of SU_2. When $c_1 \neq c_2$ and $s_{1,c_1} \neq s_{2,c_2}$, or $c_1 = c_2$ and $s_{1,c_1} = s_{2,c_2}$, there are four possible action regions, $\Psi_i, i = 0, 1, 2, 3$, for SU_2 on the plane of b_0 and p as follows.*

- $\Psi_0 = \{(b_0, p)|b_{2,i} < \underline{b}, b_{2,-i} < \underline{b}\}$ *with the optimal action* $a_2^* = 0$,
- $\Psi_1 = \bar{\Psi}_0^i \cup (\{(b_0, p)|\frac{b_{2,i}}{b_{2,-i}} > \frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G})}\} \cap \Phi_0 \cap \bar{\Psi}_0^{i,-i}) \cup (\{(b_0, p)|\frac{b_{2,i}}{b_{2,-i}} > \frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G},2)}\} \cap \bar{\Phi}_0 \cap \bar{\Psi}_0^{i,-i})$ *with the optimal action* $a_2^* = i$,
- $\Psi_2 = \bar{\Psi}_0^{-i} \cup (\{(b_0, p)|\frac{b_{2,-i}}{b_{2,i}} > \frac{R_{2,i}(\mathbf{G})}{R_{2,-i}(\mathbf{G})}\} \cap \Phi_0 \cap \bar{\Psi}_0^{i,-i}) \cup (\{(b_0, p)|\frac{b_{2,-i}}{b_{2,i}} > \frac{R_{2,i}(\mathbf{G})}{R_{2,-i}(\mathbf{G},2)}\} \cap \bar{\Phi}_0 \cap \bar{\Psi}_0^{i,-i})$ *with the optimal action* $a_2^* = -i$,
- $\Psi_3 = \Omega - \Psi_0 - \Psi_1 - \Psi_2$ *with the optimal action* $a_2^* = -a_1$,

where $-i \in \mathcal{K}\backslash i$, $\Phi_0 = \{(b_0, p)|b_{1,i} < \underline{b}, b_{1,-i} < \underline{b}\}$, $\bar{\Phi}_0 = \Omega - \Phi_0$, $\bar{\Psi}_0^i = \{(b_0, p)|b_{2,i} \geq \underline{b} > b_{2,-i}\}$, $\bar{\Psi}_0^{-i} = \{(b_0, p)|b_{2,-i} \geq \underline{b} > b_{2,i}\}$, $\bar{\Psi}_0^{i,-i} = \{(b_0, p)|b_{2,i} \geq \underline{b}, b_{2,-i} \geq \underline{b}\}$, $\Omega = \{(b_0, p)|1 > b_0 > 0, 1 > q > 0\}$, *and* $b_{1,i}$, $b_{1,-i}$, $b_{2,i}$ *and* $b_{2,-i}$ *are given by (2).*

On the other hand, when $c_1 = c_2$ and $s_{1,c_1} \neq s_{2,c_2}$, or $c_1 \neq c_2$ and $s_{1,c_1} = s_{2,c_2}$, there will be only two possible optimal actions on the whole plane of b_0 and p, that is, Ψ_0 and $\bar{\Psi}_0 = \Omega - \Psi_0$.

Proof. In case that $c_1 \neq c_2$ and $s_{1,c_1} \neq s_{2,c_2}$, or $c_1 = c_2$ and $s_{1,c_1} = s_{2,c_2}$, if (b_0, p) falls into region Ψ_0, SU_2's belief on both channels will be below the belief threshold when s_{1,c_1} and s_{2,c_2} are given, indicating SU_2 can access neither of the two channels. Otherwise, SU_2 can choose one of the channels to access. The latter case will be discuss in detail in the following.

Suppose H_i is SU_2's preferential channel. According to (4), SU_2's best response to SU_1's action a_1 can be written as follows

$$BE_2(b_0, p, \mathbf{G}, a_1 = 0, s_{1,c_1}, s_{2,c_2}) = \begin{cases} i, & \text{if } b_{2,i}R_{2,i}(\mathbf{G}) > b_{2,-i}R_{2,-i}(\mathbf{G}) \\ & \text{and } (b0, p) \in \bar{\Psi}_0^{i,-i}, \text{ or } (b0, p) \in \bar{\Psi}_0^i, \\ -i, & \text{if } b_{2,-i}R_{2,-i}(\mathbf{G}) > b_{2,i}R_{2,i}(\mathbf{G}) \\ & \text{and } (b0, p) \in \bar{\Psi}_0^{i,-i}, \text{ or } (b0, p) \in \bar{\Psi}_0^{-i}. \end{cases} \tag{37}$$

$$BE_2(b_0, p, \mathbf{G}, a_1 = i, s_{1,c_1}, s_{2,c_2}) = \begin{cases} i, & \text{if } b_{2,i}R_{2,i}(\mathbf{G},2) > b_{2,-i}R_{2,-i}(\mathbf{G}) \text{ and} \\ & (b0, p) \in \bar{\Psi}_0^{i,-i}, \text{ or } (b0, p) \in \bar{\Psi}_0^i, \\ -i, & \text{if } b_{2,-i}R_{2,-i}(\mathbf{G}) > b_{2,i}R_{2,i}(\mathbf{G},2) \\ & \text{and } (b0, p) \in \bar{\Psi}_0^{i,-i}, \text{ or } (b0, p) \in \bar{\Psi}_0^{-i}, \end{cases} \tag{38}$$

and

$$BE_2(b_0, p, \mathbf{G}, a_1 = -i, s_{1,c_1}, s_{2,c_2}) = \begin{cases} i, & \text{if } b_{2,i}R_{2,i}(\mathbf{G}) > b_{2,-i}R_{2,-i}(\mathbf{G},2) \text{ and} \\ & (b0, p) \in \bar{\Psi}_0^{i,-i}, \text{ or } (b0, p) \in \bar{\Psi}_0^i, \\ -i, & \text{if } b_{2,-i}R_{2,-i}(\mathbf{G},2) > b_{2,i}R_{2,i}(\mathbf{G}) \text{ and} \\ & (b0, p) \in \bar{\Psi}_0^{i,-i}, \text{ or } (b0, p) \in \bar{\Psi}_0^{-i}, \end{cases} \tag{39}$$

where $-i \in \mathcal{K}\backslash i$. Given s_{1,c_1} and s_{2,c_2}, SU_2's new belief on H_i and H_{-i}, i.e., $b_{2,i}$ and $b_{2,-i}$, can be calculated using (2). $\bar{\Psi}_0^i$, $\bar{\Psi}_0^{-i}$, and $\bar{\Psi}_0^{i,-i}$ are defined in Theorem 3.

Since $R_{2,i}(\mathbf{G}) > R_{2,i}(\mathbf{G},2)$ and $R_{2,-i}(\mathbf{G}) > R_{2,-i}(\mathbf{G},2)$, we have $\frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G},2)} > \frac{R_{2,-i}(\mathbf{G},2)}{R_{2,i}(\mathbf{G})}$. Moreover, $a_1 = 0$ means that (b_0, p) falls into region Φ_0. By reorganizing (37)–(39), we know that given $\mathbf{G}$, s_{1,c_1} and s_{2,c_2}, if b_0 and p fall in region Ψ_1, SU_2's optimal action is to access channel H_i, i.e., $a_2^* = i$. Similarly, if b_0 and p fall in region Ψ_2, SU_2's best response is to access H_{-i}, i.e., $a_2^* = -i$. However, if (b_0, p) lies in the region Ψ_3, the optimal action of SU_2 is to choose a different channel from SU_1, i.e., $a_2^* = -a_1$.

In case that $c_1 = c_2$ and $s_{1,c_1} \neq s_{2,c_2}$, or $c_1 \neq c_2$ and $s_{1,c_1} = s_{2,c_2}$, SU_2's belief on the two channels will be the same, i.e., $b_{2,i} = b_{2,-i}$. In such a case, there will be only two possible optimal action for SU_2, that is, Ψ_0 and $\bar{\Psi}_0 = \Omega - \Psi_0$ where SU_2's optimal action is only affected by SU_1's optimal action regardless b_0 and p. □

Appendix B. Proof of Theorem 4

Theorem 4. *Suppose H_j is the preferential channel of SU_1. Then, SU_1's optimal actions and the corresponding action regions, $\Phi_i, i = 0, 1, 2$, can be written as follows.*

- Φ_0 *with the optimal action* $a_1^* = 0$,
- $\Phi_1 = \cup_d \phi_d$ *with the optimal action* $a_1^* = j$,
- $\Phi_2 = \cup_d \bar{\phi}_d$ *with the optimal action* $a_1^* = -j$,

where $-j \in \mathcal{K}\backslash j$, $d \in \mathcal{D} = \{0, 1, 2, 3\}$, ϕ_d *and* $\bar{\phi}_d$ *are defined in (42) and (43), respectively.*

Proof. First of all, if (b_0, p) falls into region Φ_0, SU_1's belief on both channels will be below the belief threshold when s_{1,c_1} is known, which means SU_1 cannot access either of the two channels. Otherwise, SU_1 can choose one of the channels to access. In the following, we will discuss the latter case in detail.

Due to the negative network externality, SU_1's actions in $\bar{\Phi}_0$ must take into account the action of SU_2. From Theorem 3, we can see that there are four possible action regions for SU_2 in the plane of b_0 and p, i.e., Ψ_d, $d \in \mathcal{D} = \{0, 1, 2, 3\}$. In each of these regions, we can derive the optimal strategy and corresponding optimal action region for SU_1.

Let $\bar{b}_{1,j}(b_0, p, \mathbf{G}, s_{1,c_1}|\Psi_d)$ be SU_1's pseudo belief on channel H_j under region $(b_0, p) \in \Psi_d$, which is defined as follows

$$\bar{b}_{1,j}(b_0, p, \mathbf{G}, s_{1,c_1}|\Psi_d) = \sum_{\boldsymbol{\theta} \in \Theta} \sum_{x \in \{1,2\}} \alpha_{1,j}(\mathbf{G}, x) Pr(\boldsymbol{\theta}|b_0, p, s_{1,c_1}) \cdot Pr(v_{1,j} = x|(b_0, p) \in \Psi_d, \mathbf{G}, s_{1,c_1}, a_1 = j, \boldsymbol{\theta}) \mathbf{1}(\theta_j = 1), \tag{40}$$

where $\alpha_{m,k}(\mathbf{G},x)$ is the portion of maximal rate SU_m can obtain in H_k when there are x SUs sharing this channel, i.e., $\alpha_{m,k}(\mathbf{G},x)=\frac{R_{m,k}(\mathbf{G},x)}{R_{m,k}(\mathbf{G})}$.

Suppose H_j is SU_1's preferential channel, according to (5), SU_1's best response can be expressed as

$$BE_1(b_0,p,\mathbf{G},s_{1,c_1}|\Psi_d) = \begin{cases} j, & \text{if } \bar{b}_{1,j}R_{1,j}(\mathbf{G}) > \bar{b}_{1,-j}R_{1,-j}(\mathbf{G}) \text{ and } \\ & (b0,p)\in\bar{\Phi}_0^{j,-j}, \text{ or } (b0,p)\in\bar{\Phi}_0^{j}, \\ -j, & \text{if } \bar{b}_{1,-j}R_{1,-j}(\mathbf{G}) > \bar{b}_{1,j}R_{1,j}(\mathbf{G}) \text{ and } \\ & (b0,p)\in\bar{\Phi}_0^{j,-j}, \text{ or } (b0,p)\in\bar{\Phi}_0^{-j}. \end{cases} \tag{41}$$

where $-j\in\mathcal{K}\backslash j$, $\bar{\Phi}_0^{j}=\{(b_0,p)|b_{1,j}\ge \underline{b}>b_{1,-j}\}$, $\bar{\Phi}_0^{-j}=\{(b_0,p)|b_{1,-j}\ge \underline{b}>b_{1,j}\}$, and $\bar{\Phi}_0^{j,-j}=\{(b_0,p)|b_{1,j}\ge \underline{b},b_{1,-j}\ge \underline{b}\}$.

According to (41), for each Ψ_d, SU_1 will access channel H_j when (b_0,p) lies in the action region ϕ_d defined as follows:

$$\phi_d = \left(\left\{(b_0,p)|\frac{\bar{b}_{1,j}(b_0,p,\mathbf{G},s_{1,c_1}|\Psi_d)}{\bar{b}_{1,-j}(b_0,p,\mathbf{G},s_{1,c_1}|\Psi_d)} > \frac{R_{1,-j}(\mathbf{G})}{R_{1,j}(\mathbf{G})}\right\}\cap\bar{\Phi}_0^{j,-j}\right)\cup\bar{\Phi}_0^{j}, \tag{42}$$

and access channel H_{-j} when (b_0,p) lies in the action region $\bar{\phi}_d$ defined as follows:

$$\bar{\phi}_d = \left(\left\{(b_0,p)|\frac{\bar{b}_{1,j}(b_0,p,\mathbf{G},s_{1,c_1}|\Psi_d)}{\bar{b}_{1,-j}(b_0,p,\mathbf{G},s_{1,c_1}|\Psi_d)} < \frac{R_{1,-j}(\mathbf{G})}{R_{1,j}(\mathbf{G})}\right\}\cap\bar{\Phi}_0^{j,-j}\right)\cup\bar{\Phi}_0^{-j}. \tag{43}$$

By combining the results for all four action regions of SU_2, we can derive SU_1's optimal strategies and the corresponding action regions as stated in Theorem 4. □

Appendix C. Proof of Lemma 1

Lemma 1. *Given* $\mathbf{G}$, SU_2 *will choose the same optimal strategy in the action region* $\Psi_d(b_0,p)$ *with sensing results* (s_1,s_2) *and the action region* $\Psi_d(b_0,1-p)$ *with sensing results* $(\bar{s}_1,\bar{s}_2)$.

Proof. Let $\Psi_d(b_0,p)$ and $\Psi_d'(b_0,p)$, $d\in\mathcal{D}$, be SU_2's action regions with $(s_1,s_2)\in\{(s_{1,c_1},s_{2,c_2}),(\bar{s}_{1,c_1},s_{2,c_2}),(s_{1,c_1},\bar{s}_{2,c_2}),(\bar{s}_{1,c_1},\bar{s}_{2,c_2})\}$ and $(\bar{s}_1,\bar{s}_2)$, respectively. Given $\mathbf{G}$ and suppose H_i, $i\in\mathcal{K}$, is SU_2's preferential channel. According to the definition in Theorem 3, we have

$$\Psi_1'(b_0,p) = \left\{(b_0,p)|\frac{b_{2,i}(b_0,p,\bar{s}_1,\bar{s}_2)}{b_{2,-i}(b_0,p,\bar{s}_1,\bar{s}_2)} > \frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G},2)}\right\}, \tag{44}$$

$$\Psi_2'(b_0,p) = \left\{(b_0,p)|\frac{R_{2,-i}(\mathbf{G},2)}{R_{2,i}(\mathbf{G})} < \frac{b_{2,i}(b_0,p,\bar{s}_1,\bar{s}_2)}{b_{2,-i}(b_0,p,\bar{s}_1,\bar{s}_2)} < \frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G},2)}\right\}, \tag{45}$$

and

$$\Psi_3'(b_0,p) = \left\{(b_0,p)|\frac{b_{2,i}(b_0,p,\bar{s}_1,\bar{s}_2)}{b_{2,-i}(b_0,p,\bar{s}_1,\bar{s}_2)} < \frac{R_{2,-i}(\mathbf{G},2)}{R_{2,i}(\mathbf{G})}\right\}. \tag{46}$$

According to (2) and following Bayesian rule, we have

$$b_{2,i}(b_0,p,\bar{s}_1,\bar{s}_2) = b_{2,i}(b_0,1-p,s_1,s_2), \tag{47}$$

and

$$b_{2,-i}(b_0,p,\bar{s}_1,\bar{s}_2) = b_{2,-i}(b_0,1-p,s_1,s_2). \tag{48}$$

Substituting (47) and (48) into (44), (45) and (46), we have

$$\Psi_1'(b_0,p) = \left\{(b_0,p)|\frac{b_{2,i}(b_0,1-p,s_1,s_2)}{b_{2,-i}(b_0,1-p,s_1,s_2)} > \frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G},2)}\right\} = \Psi_1(b_0,1-p), \tag{49}$$

$$\Psi_2'(b_0,p) = \left\{(b_0,p)|\frac{R_{2,-i}(\mathbf{G},2)}{R_{2,i}(\mathbf{G})} < \frac{b_{2,i}(b_0,1-p,s_1,s_2)}{b_{2,-i}(b_0,1-p,s_1,s_2)} < \frac{R_{2,-i}(\mathbf{G})}{R_{2,i}(\mathbf{G},2)}\right\} = \Psi_2(b_0,1-p), \tag{50}$$

and

$$\Psi_3'(b_0,p) = \left\{(b_0,p)|\frac{b_{2,i}(b_0,1-p,s_1,s_2)}{b_{2,-i}(b_0,1-p,s_1,s_2)} < \frac{R_{2,-i}(\mathbf{G},2)}{R_{2,i}(\mathbf{G})}\right\} = \Psi_3(b_0,1-p). \tag{51}$$

Therefore, given $\mathbf{G}$, if SU_2 has action region $\Psi_d(b_0,p)$, $d\in\mathcal{D}$, with (s_1,s_2), then with $(\bar{s}_1,\bar{s}_2)$, the corresponding action region where SU_2 will adopt the same strategy is $\Psi_d(b_0,1-p)$. □

Appendix D. Proof of Lemma 2

Lemma 2. *Given* $\mathbf{G}$, SU_1 *will choose the same optimal strategy in the action region* $\phi_d(b_0,p)$ *with sensing results* s_1 *and the action region* $\phi_d(b_0,1-p)$ *with sensing results* $\bar{s}_1$.

Proof. Given $\mathbf{G}$, let $\phi_d(b_0,p)$ and $\phi_d'(b_0,p)$, $d\in\mathcal{D}$, be SU_1's action region in Ψ_d with $s_1\in\{s_{1,c_1},\bar{s}_{1,c_1}\}$ and Ψ_d' with $\bar{s}_1$, respectively. Here, Ψ_d is SU_2's action region defined in Theorem 3 with s_1 and $s_2\in\{s_{2,c_2},\bar{s}_{2,c_2}\}$ while Ψ_d' is SU_2's action region with the same strategy and $(\bar{s}_1,\bar{s}_2)$. With Lemma 1, we have $\Psi_d'(b_0,p)=\Psi_d(b_0,1-p)$.

Suppose H_j is SU_1's preferential channel, then, according to the definition of ϕ_d and ϕ_d' in Theorem 4, we have

$$\phi_d(b_0,p) = \left\{(b_0,p)|\frac{\bar{b}_{1,j}(b_0,p,\mathbf{G},s_1|\Psi_d(b_0,p))}{\bar{b}_{1,-j}(b_0,p,\mathbf{G},s_1|\Psi_d(b_0,p))} > \frac{R_{1,-j}}{R_{1,j}}\right\}, \tag{52}$$

and

$$\phi_d'(b_0,p) = \left\{(b_0,p)|\frac{\bar{b}_{1,j}(b_0,p,\mathbf{G},\bar{s}_1|\Psi_d'(b_0,p))}{\bar{b}_{1,-j}(b_0,p,\mathbf{G},\bar{s}_1|\Psi_d'(b_0,p))} > \frac{R_{1,-j}}{R_{1,j}}\right\}. \tag{53}$$

Let a_1 and a_1' be SU$_1$'s actions in regions $\Psi_d(b_0, p)$ and $\Psi_d'(b_0, p)$, respectively. According to the definition of $\Psi_d(b_0, p)$ and $\Psi_d'(b_0, p)$, for any $s_2 \in \mathcal{S}_{2,u(a_1,s_1)}$ in $\Psi_d(b_0, p)$, we have $\bar{s}_2 \in \mathcal{S}_{2,u(a_1',\bar{s}_1)}$ in $\Psi_d'(b_0, p)$. Then for $\forall x \in \{1, 2\}$ and $\forall u \in \mathcal{K}$, since SU$_2$ is the last user, given a_1 and a_2, the value of $Pr(v_{2,j} = x - 1|b_0, p, a_1 = j, a_2 = u, s_1, s_2, \boldsymbol{\theta})$ is either 0 or 1, which means it is independent with s_2 and p. Therefore, we have

$$
\begin{aligned}
&Pr(v_{1,j} = x|(b_0, p) \in \Psi_d(b_0, p), s_1, a_1 = j, \boldsymbol{\theta}) \\
&= \sum_{u\in\mathcal{K}} \sum_{s_2\in\mathcal{S}_{2,u(a_1,s_1)}} Pr(v_{2,j} = x - 1|(b_0, p) \in \Psi_d(b_0, p), \\
&\quad a_1 = j, a_2 = u, s_1, s_2, \boldsymbol{\theta}) f(s_2|\boldsymbol{\theta}) \\
&= \sum_{u\in\mathcal{K}} Pr(v_{2,j} = x - 1|(b_0, p) \in \Psi_d(b_0, p), a_1 = j, \\
&\quad a_2 = u, s_1, s_2, \boldsymbol{\theta}) \sum_{s_2\in\mathcal{S}_{2,u(a_1,s_1)}} f(s_2|\boldsymbol{\theta}) \\
&= y(p),
\end{aligned} \tag{54}
$$

where $y(p)$ is defined as a function of p.

Similarly, we have

$$
\begin{aligned}
&Pr(v_{1,j} = x|(b_0, p) \in \Psi_d'(b_0, p), \bar{s}_1, a_1' = j, \boldsymbol{\theta}) \\
&= \sum_{u\in\mathcal{K}} Pr(v_{2,j} = x - 1|(b_0, p) \in \Psi_d'(b_0, p), \\
&\quad a_1' = j, a_2 = u, \bar{s}_1, \bar{s}_2, \boldsymbol{\theta}) \sum_{\bar{s}_2\in\mathcal{S}_{2,u(a_1',\bar{s}_1)}} f(\bar{s}_2|\boldsymbol{\theta}) \\
&= y(1 - p).
\end{aligned} \tag{55}
$$

Moreover, $\forall \boldsymbol{\theta} \in \Theta$, let $Pr(\boldsymbol{\theta}|s_1) = g(p)$, we have $Pr(\boldsymbol{\theta}|\bar{s}_1) = g(1 - p)$. Then, according to (40), (54), (55) and Lemma 1, we have

$$
\begin{aligned}
&\bar{b}_{1,j}(b_0, p, \mathbf{G}, \bar{s}_1|\Psi_d'(b_0, p)) \\
&= \bar{b}_{1,j}(b_0, 1 - p, \mathbf{G}, s_1|\Psi_d(b_0, 1 - p)),
\end{aligned} \tag{56}
$$

and

$$
\begin{aligned}
&\bar{b}_{1,-j}(b_0, p, \mathbf{G}, \bar{s}_1|\Psi_d'(b_0, p)) \\
&= \bar{b}_{1,-j}(b_0, 1 - p, \mathbf{G}, s_1|\Psi_d(b_0, 1 - p)).
\end{aligned} \tag{57}
$$

By substituting (56) and (57) into (52) and (53), we have

$$
\phi_d'(b_0, p) = \phi_d(b_0, 1 - p). \tag{58}
$$

□

References

[1] B. Zhang, Y. Chen, C. Wang, K.J.R. Liu, Learning and decision making with negative externality for opportunistic spectrum access, in: Proceedings of IEEE Global Communications Conference, GLOBECOM, 2012.

[2] S. Haykin, D.J. Thomson, J.H. Reed, Spectrum sensing for cognitive radio, Proc. IEEE 97 (5) (2009) 849–877.

[3] K. Hyoil, K.G. Shin, Optimal online sensing sequence in multichannel cognitive radio network, IEEE Trans. Mobile Comput. 12 (7) (2013) 1349–1362.

[4] L. Ma, X. Han, C.C. Shen, Dynamic open spectrum sharing MAC protocol for wireless ad hoc networks, in: Proceedings of IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005.

[5] R. Urgaonkar, M.J. Neely, Opportunistic scheduling with reliability guarantees in cognitive radio networks, IEEE Trans. Mobile Comput. 8 (6) (2009) 766–777.

[6] Q. Zhao, L. Tong, A. Swami, Y. Chen, Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework, IEEE J. Sel. Areas Commun. 25 (3) (2007) 589–600.

[7] J. Park, P. Paweczak, D. Cabric, Performance of joint spectrum sensing and MAC algorithms for multichannel opportunistic spectrum access ad hoc networks, IEEE Trans. Mobile Comput. 10 (7) (2011) 1011–1027.

[8] K.W. Choi, E. Hossain, Opportunistic access to spectrum holes between packet bursts: a learning-based approach, IEEE Trans. Wirel. Commun. 10 (8) (2011) 2497–2509.

[9] V. Kone, L. Yang, X. Yang, B.Y. Zhao, H. Zheng, On the feasibility of effective opportunistic spectrum access, in: Proceedings of Internet Measurement Conference, IMC, 2010.

[10] T.M. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.ISBN 0070428077

[11] B. Wang, Y. Wu, K.J.R. Liu, Anti-jamming stochastic game in cognitive radio networks, IEEE J. Sel. Areas Commun. 29 (4) (2011) 877–889.Special issue on Cognitive Radio Networking and Communications

[12] J. Lunden, V. Koivunen, A. Huttunen, H.V. Poor, Spectrum sensing in cognitive radios based on multiple cyclic frequencies, in: Proceedings of IEEE Crowncom, Orlando, Florida, USA, 2007.

[13] M. Gandetto, C. Regazzoni, Spectrum sensing: a distributed approach for cognitive terminals, IEEE J. Sel. Areas Commun. 25 (3) (2007) 546–557.

[14] B. Golub, M.O. Jackson, Naive learning in social networks and the wisdom of crowds, Am. Econ. J.: Microecon. 2 (1) (2010) 112–149.

[15] H. Carlsson, E.V. Damme, Global games and equilibrium selection, Econom.: J. Econom. Soc. 61 (5) (1993) 989–1018.

[16] B. Vcking, R. Aachen, Congestion games: optimization in competition, in: Proceedings of the 2nd Algorithms Complexity Durham Workshop, Kings College Publications, 2006, pp. 9–20.

[17] C. Wang, Y. Chen, K.J.R. Liu, Sequential Chinese restaurant game, IEEE Trans. Signal Process. 61 (3) (2013) 571–584.

[18] C. Wang, Y. Chen, K.J.R. Liu, Chinese restaurant game, IEEE Signal Process. Lett. 19 (12) (2012) 898–901.

[19] Q. Zhao, B.M. Sadler, A survey of dynamic spectrum access, IEEE Signal Process. Mag. 24 (3) (2007) 79–89.

[20] K. Liu, Q. Zhao, A restless bandit formulation of opportunistic access: indexablity and index policy, in: Proceedings of IEEE Conference on Sensor, Mesh and Ad Hoc Comm. and Networks, SECON, 2008.

[21] A. Anandkumar, N. Michael, A. Tang, Opportunistic spectrum access with multiple users: learning under competition, in: Proceedings of IEEE Conference on Computer Communications, INFOCOM, San Deigo, USA, 2010.

[22] A. Anandkumar, N. Michael, A. Tang, A. Swami, Distributed algorithms for learning and cognitive medium access with logarithmic regret, IEEE J. Sel. Areas Commun. 29 (4) (2011) 731–745.

[23] X. Li, P. Yang, Y. Yan, L. You, S. Tang, Q. Huang, Almost optimal accessing of nonstochastic channels in cognitive radio networks, in: Proceedings of IEEE Conference on Computer Communications, INFOCOM, 2012, pp. 2291–2299.

[24] V. Krishnamurthy, Decentralized spectrum access amongst cognitive radios: an interacting multivariate global game-theoretic approach, IEEE Trans. Signal Process. 57 (10) (2009) 3999–4013.

[25] L.M. Law, J. Huang, M. Liu, Price of anarchy for cogestion game in cogonitive radio networks, IEEE Trans. Wirel. Commun. 11 (10) (2012) 3778–3787.

[26] G.M. Angeletos, C. Hellwig, A. Pavan, Signaling in a global game: coordination and policy traps, J. Political Econ. 114 (3) (2006) 452–484.

[27] G.M. Angeletos, C. Hellwig, A. Pavan, Dynamic global games of regime change: learning, multiplicity, and the timing of attacks, Econometrica 75 (3) (2007) 711–756.

[28] R. Kleinberg, G. Piliouras, E. Tardos, Multiplicative updates outperform generic no-regret learning in congestion games, Proceedings of ACM Symposium on theory of computing, STOC, Bethesda, MD, 2009.

[29] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, Y.-D. Yao, Opportunistic spectrum access in unknown dynamic environment: a game-theoretic stochastic learning solution, IEEE Trans. Wirel. Commun. 11 (4) (2012) 1380–1391.

[30] C. Jiang, Y. Chen, Y. Yang, C. Wang, K.J.R. Liu, Dynamic Chinese restaurant game: theory and application to cognitive radio networks, IEEE Trans. Wirel. Commun. 13 (4) (2014) 1960–1973.

[31] N. Lee, J. X. Lin, G. Andrews, R.W. Heath, Power control for D2D underlaid cellular networks: modeling, algorithms, and analysis, IEEE J. Sel. Areas Commun. 33 (1) (2015) 1–13.

[32] A. Sakr, E. Hossain, Cognitive and energy harvesting-based D2D communication in cellular networks: Stochastic geometry modeling and analysis, IEEE Trans. Wirel. Commun. 63 (5) (2015) 1867–1880.

[33] H. Yao, Z. Zhou, H. Liu, L. Zhang, Optimal power allocation in joint spectrum underlay and overlay cognitive radio networks cognitive radio oriented wireless networks and communications, in: Proceedings of IEEE Cognitive Radio Oriented Wireless Networks and Communications, CROWNCOM, 2009, pp. 1–5.
[34] A.T. Hoang, Y.-C. Liang, Downlink channel assignment and power control for cognitive radio networks, IEEE Trans. Wirel. Commun. 7 (8) (2008) 3106–3117.
[35] Z. Han, R. Zheng, H.V. Poor, Repeated auctions with Bayesian nonparametric learning for spectrum access in cognitive radio networks, IEEE Trans. Wirel. Commun. 10 (3) (2011) 890–900.
[36] Y. Chen, Q. Zhao, A. Swami, Distributed spectrum sensing and access in cognitive radio networks with energy constraint, IEEE Trans. Signal Process. 57 (2) (2009) 783–797.
[37] M.J. Osborne, A. Rubenstein, A Course in Game Theory, MIT Press, 1994.
[38] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, R. Pickholtz, Spectrum leasing to cooperating secondary ad hoc networks, IEEE J. Sel. Areas Commun. 26 (1) (2008) 203–213.
[39] Y. Yi, J. Zhang, Q. Zhang, T. Jiang, Spectrum leasing to multiple cooperating secondary cellular networks, in: Proceedings of IEEE International Conference on Communications, ICC, 2011.
[40] J. Yu, B. Zhang, P. Chen, Blind and semi-blind channel estimation with fast convergence for MIMO-OFDM systems, Signal Process. 95 (2014) 1–9.
[41] B. Zhang, Y. Chen, C. Wang, K.J.R. Liu, Supplementary Information For a Chinese Restaurant Game for Learning and Decision Making in Cognitive Radio Networks', 2014, Online at http://sig.umd.edu/yan/CRG_Biling.pdf.
[42] M.D. Felice, K.R. Chowdhury, L. Bononi, Learning with the bandit: a cooperative spectrum selection scheme for cognitive radio networks, in: Proceedings of IEEE Global Communications Conference, GLOBECOM, 2011.

Biling Zhang (M'12) received the B.S., M.S. and Ph.D degrees from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2000, 2003 and 2013, respectively. Now she is an Assistant Professor in BUPT. From February 2011 to March 2012, she has been a visiting scholar in the Department of Electrical and Computer Engineering at the University of Maryland, College Park, MD, USA. Her current research interests are in the areas of cognitive radio networks, wireless communication and cooperative communications.

Yan Chen (S'06-M'11) received the Bachelor's degree from University of Science and Technology of China in 2004, the M.Phil degree from Hong Kong University of Science and Technology (HKUST) in 2007, and the Ph.D. degree from University of Maryland College Park in 2011. He is currently a research associate in the Department of Electrical and Computer Engineering at University of Maryland College Park. His current research interests are in social learning and networking, smart grid, cloud computing, crowdsourcing, network economics, multimedia signal processing and communication. Dr. Chen received the University of Maryland Future Faculty Fellowship in 2010, Chinese Government Award for outstanding students abroad in 2011, University of Maryland ECE Distinguished Dissertation Fellowship Honorable Mention in 2011, and was the Finalist of A. James Clark School of Engineering Dean's Doctoral Research Award in 2011.

Chih-Yu Wang received the B.S. degree in Electrical Engineering from National Taiwan University. in 2007. He received the Ph.D. degree from Graduate Institute of Communication Engineering, National Taiwan University. He has been a visiting student in University of Maryland, College Park in 2011. He is currently an Assistant Research Fellow at the Research Center for Information Technology Innovation, Academia Sinica. His research interests include game theory, wireless communications, social networks, and data science.

K. J. Ray Liu (F'03) was named a Distinguished Scholar-Teacher of University of Maryland, College Park, in 2007, where he is Christine Kim Eminent Professor of Information Technology. He leads the Maryland Signals and Information Group conducting research encompassing broad areas of signal processing and communications with recent focus on cooperative communications, cognitive networking, social learning and networks, and information forensics and security. Dr. Liu is the recipient of numerous honors and awards including IEEE Signal Processing Society Technical Achievement Award and Distinguished Lecturer. He also received various teaching and research recognitions from University of Maryland including university-level Invention of the Year Award; and Poole and Kent Senior Faculty Teaching Award and Outstanding Faculty Research Award, both from A. James Clark School of Engineering. An ISI Highly Cited Author, Dr. Liu is a Fellow of IEEE and AAAS. Dr. Liu is President of IEEE Signal Processing Society where he has served as Vice President - Publications and Board of Governor. He was the Editor-in-Chief of IEEE Signal Processing Magazine and the founding Editor-in-Chief of EURASIP Journal on Advances in Signal Processing.