Robust Multiuser Sequential Channel Sensing and Access in Dynamic Cognitive Radio Networks: Potential Games and Stochastic Learning

Yuhua Xu, Qihui Wu, *Senior Member, IEEE*, Jinlong Wang, *Senior Member, IEEE*, Liang Shen, and Alagan Anpalagan, *Senior Member, IEEE*

***Abstract*—This paper studies the problem of multiuser sequential channel sensing and access in dynamic cognitive radio networks, in which the active-user set is randomly changing from slot to slot. Furthermore, each user only has its individual information with no information exchange among users. The goal of the users is to determine their channel sensing order. We first define a generalized interference metric to address the overlapping of channel sensing order and establish two optimization objectives: minimizing the aggregate interference for each active-user set and minimizing the expected aggregate interference for all potential users. It is challenging to solve the two optimization problems, even in a centralized manner, because the active-user set is randomly changing, and the probability distributions of the active-user sets are unknown to the users. We then propose two noncooperative game models to solve the optimization problems: a state-based one-shot game and a robust game. We prove that they are potential games and that the best Nash equilibrium of the two games corresponds to the optimal solutions of the two optimization problems, respectively. To cope with the *uncertain*, *dynamic*, and *incomplete* information constraints in the dynamic networks, we propose a stochastic learning algorithm, which is analytically proven to converge to Nash equilibria of the two formulated games in the presence of a changing active-player set. Finally, simulation results are presented to validate the convergence and superior performance of the proposed learning algorithm.**

***Index Terms*—Cognitive radio (CR) networks, multiuser stochastic learning, noncooperative game, potential game, sequential channel sensing and access.**

I. Introduction

COGNITIVE radio (CR), which was first coined by Mitola and Maguire in [1], has been regarded as a promising approach to lessen the dilemma between spectrum shortage and spectrum waste [2]–[7]. Due to hardware limitations, secondary users (SUs) can only sense a part of the licensed channels (always one) at a time [8]. As a result, SUs sense the licensed channels one by one according to a predefined order, which is referred to as the sequential sensing strategy [9]. In this setting, collision occurs if more than two SUs sense and access an idle channel simultaneously, which implies that the sensing orders should be carefully designed [10]–[13]. In this paper, we focus on the problem of channel sensing order optimization in multiuser CR networks.

In multiuser CR networks, some learning approaches are needed to coordinate the behaviors of SUs [14]. Due to the dynamic spectrum opportunities and limited information, it generally takes multiple slots for the learning approaches to converge to some stable solutions. The problem of the sequential channel sensing strategy for single-user CR systems has been extensively studied [15]–[21], whereas for multiuser CR networks, it has begun drawing attention [10]–[13]. In most existing work for multiuser CR networks, the number of active SUs is assumed fixed during the whole learning process. However, in several practical scenarios, an SU does not perform learning when it has no data to transmit. Thus, it is important and timely to reinvestigate this problem in the presence of a dynamic set of active users.

In this paper, we consider a distributed dynamic CR networks in which each user only has its individual information, and there is no information exchange among users. Furthermore, each user is active with a certain probability in each slot, which can be regarded as an abstraction of dynamic traffic. The goal of the users is to determine its channel sensing order. We first define a generalized interference metric to address the overlapping of multiple channel orders. Based on this, we define two optimization objectives: minimizing the aggregate interference for each active-user set and minimizing the expected aggregate interference for all potential users. Information is key to decision-making problems [22], and the challenges related to information arising in the considered dynamic network are listed as follows.

- **Uncertain**: The occupancy states of the licensed channels and the set of active users in each slot are random.
- **Dynamic**: The occupancy states of the licensed channels change from slot to slot, and the set of active users also changes from slot to slot.
- **Incomplete**: Each SU has no information of other SUs and only has partial information about the environment. Specifically, one SU does not know the active probabilities

Manuscript received March 22, 2014; revised July 26, 2014; accepted September 6, 2014. Date of publication September 9, 2014; date of current version August 11, 2015. This work was supported by the National Science Foundation of China under Grant 61401508 and Grant 61172062 and in part by the Jiangsu Province Natural Science Foundation under Grant BK2011116. The review of this paper was coordinated by Prof. J. Tang.

Y. Xu, Q. Wu, J. Wang, and L. Shen are with the College of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China (e-mail: yuhuaenator@gmail.com; wqhqhw@163.com; wjl543@sina.com; ShenLiang671104@sina.com).

A. Anpalagan is with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: alagan@ee.ryerson.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2014.2356554

of all other users, and the licensed channel idle probabilities are also unknown to the SUs.

For presentation, we denote the given features for decision-making problems as *uncertain*, *dynamic*, and *incomplete* (UDI) information constraints. It is impossible to solve the two interference minimization problems even in a centralized manner, since the active-user set is randomly changing from slot to slot, and the probability distributions of the active-user sets are unknown to the users. In this paper, we will resort to distributed learning approaches based on game theory, which is a powerful tool for the multiuser decision-making problem.

We propose two noncooperative game models to solve the given optimization problems. The first is a state-based one-shot game, in which an inherent system state describes the active-user set in each slot. Specifically, the set of active users (players[1]) in each slot is randomly determined by the active probabilities of the potential users. The second is a robust game [23], in which the utility functions are defined as the expected value over all system states. We prove that the best Nash equilibrium (NE) of the two games corresponds to the optimal solutions of the two optimization problems, respectively. However, due to the UDI information constraints, existing game-theoretic algorithms cannot be applied into the considered dynamic CR networks. We then propose a stochastic online learning algorithm that converges to the best NE of the two games under the UDI information constraints.

The main contributions of this paper are summarized as follows.

1) For distributed multiuser decision-making problems with UDI information constraints, we establish an efficient game-theoretic framework. It can be applied to several scenarios wtih other kinds of UDI information constraints, e.g., changing channel rate due to instantaneous channel fading and user mobility.
2) To capture the interactions among multiple users, we define a generalized interference metric for the channel sensing order and establish two interference minimization problems. We formulate the optimization problems as two noncooperative games (a state-based one-shot game and a robust game). We prove that the best NE of the games corresponds to the best solutions for the original optimization problems, respectively.
3) To deal with the UDI information constraints, which are mainly caused by the randomly changing active-user set, we propose a distributed stochastic learning algorithm and prove its convergence. Moreover, it is analytically proved that the proposed algorithm asymptotically converges to the best solutions of the two interference minimization problems.

The rest of this paper is organized as follows. In Section II, we give a brief review of related work. In Section III, we present the considered dynamic CR network model, define a generalized interference metric, and establish two interference mitigation problems. In Section IV, we formulate the state-based one-shot game and a robust game and investigate the properties of their Nash equilibria. In Section V, we propose a distributed stochastic learning algorithm for achieving Nash equilibria of the games in the presence of a changing active-user set. In Section VI, simulation results and discussion are presented. Finally, we present discussions and draw conclusions in Section VII.

[1]In the rest of this paper, we will use users and players interchangeably.

II. Related Work

The problem of channel sensing and access is a current research topic in CR networks. A large number of decision-making approaches were proposed in the literature, e.g., partially observable Markovian decision process [8], multiarmed bandit problems [24], and game-theoretic multiagent learning [25]–[29]. A common drawback of these approaches is that they employ the parallel sensing strategy, which may fail to find more spectrum opportunities. In the parallel sensing strategy, the SUs sense a fixed number (always one) of licensed channels at the beginning of each slot, access the idle channels, or suspend their transmissions until the next slot if the channel is detected as occupied. It is shown that the parallel sensing strategy admits analytical tractability but leads to relatively lower spectrum utilization [9]. Compared with the parallel sensing strategy, the sequential sensing strategy provides more efficient and adaptive spectrum opportunity discovery.

The sequential channel sensing and access has been extensively investigated using the optimal stopping problem (OSP) models for single-user CR systems [6], [15]–[21]. In addition, some preliminary results based on OSP models for multiuser CR systems were reported in the literature. Specifically, a centralized solution for a two-user CR system was proposed in [30], and a heuristic solution was proposed in [31].

Recently, the problem of channel sensing order optimization for multiuser CR networks, which is exactly the research topic in this paper, has begun drawing attention. In particular, an adaptive persistent sensing order selection strategy was proposed in [10], a dynamic-programming-based order selection strategy in [11], and a reinforcement-learning-based order selection algorithm in [12]. The main difference in our work is that the active-user set in each slot is randomly changing, and all the aforementioned algorithms do not converge in the presence of a changing active-user set. Moreover, a modified p-persistent access scheme for distributed multiuser sequential channel sensing in multichannel CR networks was studied in [13].

The game model is a powerful tool in investigating the interactions among multiple users and has been extensively used in wireless communication networks, e.g., distributed interference mitigation [32], power control [33], multiple user access [34], spectrum allocation [25], routing [35], and heterogeneous network selection [36]. In the methodology, almost all game models in the literature are with a fixed active-user set, whereas we consider a changing active-user set in this paper. In addition, the proposed learning algorithm is carefully designed to deal with a randomly changing active-user set in each slot. Recently, it should be pointed out that some useful tutorials for robust games with a changing player set can be found in [23].

The proposed stochastic learning algorithm is based on [49]. Due to its computational efficiency and simple implementation,

this kind of stochastic learning algorithm has been extensively used in wireless applications, e.g., rate adaptation for IEEE 802.11 networks [37], discrete power control [38], and dynamic spectrum access with quality of service and interference temperature constraints [39]. Among the existing work, the convergence for a two-user scenario was studied in [38], and the convergence for a type of game in which the same received payoffs for all the players was investigated in [39], respectively. Compared with the existing work, the new technical contributions are as follows: 1) The formulated game involves multiple users with different received payoffs, and 2) the set of active players is time varying.

III. System Model and Problem Formulation

A. System Model

Consider a distributed CR network consisting of N SUs and M licensed channels. The licensed channels are owned by the primary users (PUs) and can only be opportunistically used by the SUs when they are not occupied by the PUs. Time is divided into slots with equal length, and the activities of the PUs are slotted. Denote x_m as the occupancy state of channel m; specifically, $x_m = 1$ means that channel m is idle, and $x_m = 0$ means that it is occupied by PUs. Assume that the states of all channels remain unchanged in a slot and change randomly and independently from slot to slot and from channel to channel. Thus, we characterize the dynamic of spectrum opportunities by the idle channel probabilities, which are denoted as θ_m, $0 \leq \theta_m \leq 1 \ \forall m \in \{1, 2, \ldots, M\}$. In this paper, we assume that the number of the SUs is not greater than that of the licensed channels,[2] i.e., $N \leq M$.

Denote the slot length as T and the time for sensing one channel is T_s. The sensing performance is characterized by the detection probability $P_d(T_s)$, i.e., the probability of the event that the channel is detected as occupied and it is truly occupied, and the false alarm probability $P_f(T_s)$, i.e., the probability of the event that the channel is idle while it is detected as occupied. Analytical investigations on the two probabilities can be found in [40].

We consider a dynamic CR network, which is mainly characterized by a variable number of active users. Specifically, each SU performs channel sensing and access in each slot with probability λ_n, $0 < \lambda_n \leq 1$. Note that such a dynamic model captures a general kind of dynamics in CR networks, e.g., an SU is active only when it has data to transmit and becomes inactive when there is no transmission demand, a mobile user joins or leaves the network dynamically. Moreover, it can be regarded as an abstraction of the user traffic, i.e., the user active probability corresponds to the probability of nonempty buffer. In the considered dynamic network, the key decision of the (active) SUs is to determine their channel sensing orders distributively and autonomously.

[2]For the case of $N > M$, the SUs would choose a fixed number of channels (always one) to sense and access in a slot rather than performing the sequential channel sensing and access strategies. The inherent reason is that the resource is limited in this case.

B. Problem Formulation

To capture the variable number of active users in the considered dynamic CR network, we define the underlying system state as $\mathbf{S} = \{s_1, \ldots, s_N\}$, where $s_n = 1$ indicates that the nth SU is active, whereas $s_n = 0$ indicates that it is inactive. The probability of an underlying system state is given by $\mu(s_1, \ldots, s_N) = \prod_{n=1}^{N} p_n$, where p_n is determined as follows:

$$p_n = \begin{cases} \lambda_n, & s_n = 1 \\ 1 - \lambda_n, & s_n = 0. \end{cases} \tag{1}$$

However, it is emphasized that the state distribution probabilities are unknown to the SUs since each SU does not know the active probabilities of other SUs. Denote the potential user set as $\mathcal{N}$, i.e., $\mathcal{N} = \{1, \ldots, N\}$, and the active-user set as $\mathcal{B}$, i.e., $\mathcal{B} = \{n \in \mathcal{N} : s_n = 1\}$. For presentation, denote the set of all the active-user sets as Γ. Then, the probability of an active-user set can be given by $\mu(\mathcal{B})$, which satisfies $\sum_{\mathcal{B} \in \Gamma} \mu(\mathcal{B}) = 1$.

Due to interactions among SUs, the channel sensing order profile of all the SUs has a great impact on the achievable network throughput. For example, suppose that there is a CR system with five licensed channels and two active SUs whose channel sensing orders are given by {1, 3, 2, 4, 5} and {3, 4, 2, 5, 1}, respectively. It is seen that the two SUs simultaneously sense channel 2 at time 3τ, where τ is the sensing duration on a channel. In this configuration, if channel 2 is detected as idle by both users, they access the channel simultaneously and cause collision.[3] An illustrative diagram of the example system is shown in Fig. 1.

Denote the permutation set of M as $\mathcal{O}$. For presentation, we denote the channel sensing order of the nth SU as an M-dimensional order vector $\mathbf{O}_n = (o_{n1}, o_{n2}, \ldots, o_{nM})$, which corresponds to a permutation chosen from $\mathcal{O}$. To characterize the interactions among the SUs, we define a generalized interference metrics below. First, the generalized interference between two active users n and m is defined as follows:

$$g_{nm} = \mathbf{O}_n \odot \mathbf{O}_m \tag{2}$$

where $\odot$ is the bitwise XNOR operation. In other words, it is calculated by $g_{nm} = \sum_{k=1}^{M} \delta(o_{nk}, o_{mk})$, where $\delta(o_{nk}, o_{mk})$ is the following indicator function:

$$\delta(o_{nk}, o_{mk}) = \begin{cases} 1, & o_{nk} = o_{mk} \\ 0, & o_{nk} \neq o_{mk}. \end{cases} \tag{3}$$

Intuitively, g_{nm} reflects the impact of overlapped channel sensing orders on the achievable throughput of the two users. Specifically, a larger value of g_{nm} implies lower individual achievable throughput of them, and *vice versa*. In particular, if the interference level between n and m is zero, i.e., $g_{nm} = 0$, we say that they are *orthogonal*. Since there are multiple active

[3]The CR users can employ some resolution approaches, e.g., cognitive MAC proposed in [41], to avoid collision. However, as will be shown later, the order selection converging profile of the users is interference-profile, and they can transmit whenever an idle channel is detected, which will save time and energy overhead for collision resolution.

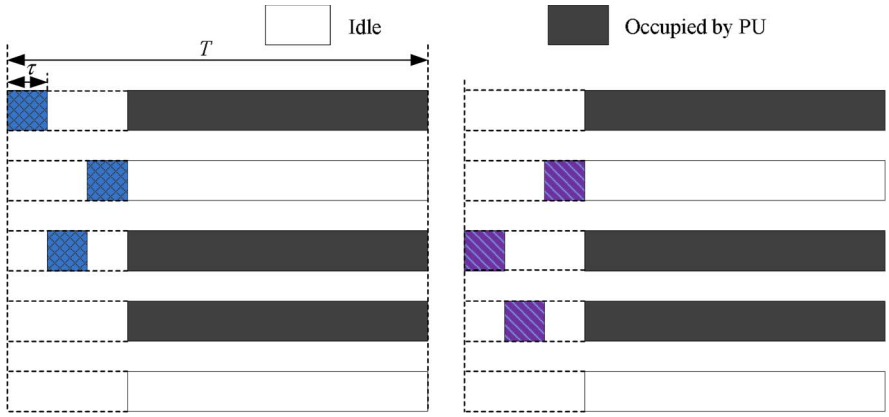

Fig. 1. CR system with five licensed channels and two SUs. SU 1's channel sensing order is {1, 3, 2, 4, 5}, and SU 2's sensing order is {3, 4, 2, 5, 1}. When the two users both detected channel 2 as idle in time 3τ, where τ is the sensing duration on a channel, they access this channel simultaneously and cause collision.

users in the system, we define the aggregate generalized level experienced by an active user n in slot k as follows:

$$G_n(\mathcal{B}) = \sum_{m\in\mathcal{B},m\neq n} g_{nm} = \sum_{m\in\mathcal{B},m\neq n} \mathbf{O}_n \odot \mathbf{O}_m. \quad (4)$$

From the user side, a lower value of $G_n(\mathcal{B})$ is desirable since it is expected to achieve higher throughput. From the system side, a lower value of the aggregate generalized interference level of all the users in the system is also desirable. This motivates us to define the aggregate generalized interference level of an active-user set $\mathcal{B}(k)$ as follows:

$$I(\mathcal{B}) = \sum_{n\in\mathcal{B}} \sum_{m\in\mathcal{B},m\neq n} \mathbf{O}_n \odot \mathbf{O}_m. \quad (5)$$

If the aggregate generalized interference level of an active-user set is zero, i.e., $I(\mathcal{B}) = 0$, it means that there is no overlap in the channel sensing orders of any two users. In other words, the channel sensing order profile is *interference free*. Clearly, an interference-free channel sensing order profile is desirable since it would lead to high network throughput, as can be expected. Mathematically, the optimization objective is to find the optimal channel sensing order profile such that the following objectives are maximized:

$$(\mathbf{P1}:) \quad \max -I(\mathcal{B}) \qquad \forall\, \mathcal{B} \in \Gamma \quad (6)$$

or

$$(\mathbf{P2}:) \quad \max -\mathrm{E}_{\mathcal{B}}\left[I(\mathcal{B})\right] = \sum_{\mathcal{B}\in\Gamma} \mu(\mathcal{B}) I_{\mathcal{B}} \quad (7)$$

where $\mathrm{E}_{\mathcal{B}}[\cdot]$ is the operation of taking expectation over all possible active-user sets. It is seen that problem $\mathbf{P1}$ is for every possible active-user set in a slot, whereas problem $\mathbf{P2}$ is in the sense of expectation. Generally, the tasks of solving problems $\mathbf{P1}$ and $\mathbf{P2}$ are challenging, since they are combinatorial optimization problems. Inherently, the active-user set in each slot is unknown to the users (a user only knows its state and does not know the states of others), which furthermore adds difficulties into solving problem $\mathbf{P1}$. For problem $\mathbf{P2}$, the distribution probabilities $\mu(\mathcal{B})$ are unknown. In fact, due to the incomplete information, $\mathbf{P1}$ and $\mathbf{P2}$ cannot be solved, even in a centralized manner. In the following, we will propose a game-theoretic learning framework for solving the two problems.

Remark 1: The optimality of the formulated optimization objectives is discussed as follows. If all the channels are homogeneous, the optimality in interference is also the optimality in throughput if the final results are interference free. On the other side, this is not true if the channels are heterogeneous. However, due to the extremely heavy complexity, it is hard to achieve the optimal sum-rate maximization solution. For example, for a system with five channels and five SUs, the possible simulations are $(5!)^5$, which is extremely huge. Thus, we think that the formulated optimization objective is desirable but not always optimal. In particular, if the final result is an interference-free solution, the throughput would be satisfactory as can be expected.

IV. Channel Sensing Order Selection Games

The fact that the selections of channel sensing orders of active users are interactive motivates us to formulate this problem as a noncooperative game. Here, we formulate two game models to capture the interactions among active users. The first game is a state-based order selection game, in which an inherent system state describes the active-user set in each slot. The second game is a robust order selection game, in which the utility functions are defined as the expected value over all system states. We analyze their properties including the existence of Nash equilibria and their achievable performance. More importantly, we quantitatively study some relationships between the steady states of the two games and the solutions of the original optimization problems $\mathbf{P1}$ and $\mathbf{P2}$, respectively.

A. State-Based Order Selection Game

1) Game Model: To address the changing active-user set, a system state is added into the game model. Specifically, the state-based one-shot game is denoted as $\mathcal{G}_1 = \{\mathcal{N}, \mathcal{B}, \{\mathcal{A}_n\}_{n\in\mathcal{B}}, \{U1_n\}_{n\in\mathcal{B}}\}$, where $\mathcal{N}$ is the potential player (SU) set, i.e., $\mathcal{N} = \{1, 2, \ldots, N\}$, $\mathcal{B}$ is the underlying system state that determines the current active-user set, $\mathcal{A}_n$ is the available action (channel sensing and access order) set of active player n, and $U1_n$ is the utility function of player n.

It is assumed that all the players' action sets are the permutation set of M, i.e., $\mathcal{A}_n = \mathcal{O},\ \forall\, n \in \mathcal{N}$. That is, each player can arbitrarily determine its sensing and access order. Denote $\mathbf{O}_n \in \mathcal{A}_n$ as an action chosen by player n. Since the actions of the players are interactive, the utility function is denoted as $U1_n(\mathbf{O}_n, \mathbf{O}_{-n})$, where $\mathbf{O}_n \in \mathcal{A}_n$ is the chosen action of player n, and $\mathbf{O}_{-n}$ is the action profile of all the active players except player n. For an inactive player, the utility function is zero; for an active player $n \in \mathcal{B}$, the utility function is defined as follows:

$$U1_n(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n}) = -\sum_{m\in\mathcal{B}, m\neq n} \mathbf{O}_n \odot \mathbf{O}_m. \tag{8}$$

In noncooperative game models, each active player intends to maximize its individual utility function [42], which means that the state-based channel sensing order selection game can be expressed as

$$\mathcal{G}_1: \quad \max U1_n(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n}) \qquad \forall\, n \in \mathcal{B}. \tag{9}$$

With problem $\mathbf{P1}$ now formulated as a noncooperative game, two questions naturally arise: 1) What are the properties of the game, e.g., the existence of Nash equilibria and their achievable performance; and 2) what are the relationships between $\mathcal{G}_1$ and the original optimization problem $\mathbf{P1}$? We will answer the two questions in the following.

2) Analysis of the NE: First, we present the definition of NE in the state-based one-shot game, which can be regarded as a generalization of NE in the traditional static games with a fixed number of players. For an arbitrary system state $\mathcal{B}$, a channel sensing order profile of all the active players $a_{\mathrm{NE}} = \{\mathbf{O}_n^*, \mathbf{O}_{-n}^*\}$ is a pure strategy NE of $\mathcal{G}_1$ if and only if no player can improve its utility function by unilaterally deviating, i.e.,

$$U1_n\left(\mathcal{B}, \mathbf{O}_n^*, \mathbf{O}_{-n}^*\right) \geq U1_n\left(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n}^*\right) \qquad \forall\, \mathbf{O}_n \in \mathcal{O} \qquad \forall\, n \in \mathcal{B}. \tag{10}$$

In addition, the aggregate interference level of all the active players in a pure strategy NE of $\mathcal{G}_1$ is given by

$$I_{\mathcal{G}_1} = -\sum_{n\in\mathcal{B}} U1_n\left(\mathcal{B}, \mathbf{O}_n^*, \mathbf{O}_{-n}^*\right). \tag{11}$$

The properties of the state-based channel sensing order selection game $\mathcal{G}_1$ is characterized by the following theorem.

Theorem 1: For any active-user set, i.e., $\forall\, \mathcal{B} \in \Gamma$, the state-based channel sensing order selection game $\mathcal{G}_1$ is an exact potential game that has at least one pure strategy NE point. Furthermore, any optimal solution of problem $\mathbf{P1}$ constitutes a pure strategy NE of the game.

Proof: To prove this theorem, we need to prove that there exists a potential function such that the change in the utility function of an active player by its unilaterally deviating is the same as that in the potential function. Specifically, we define the following state-based potential function $\Phi 1 : \mathbf{O}_n \times \mathbf{O}_{-n} \to R$ for the formulated game:

$$\Phi 1(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n}) = -\frac{1}{2}\sum_{n\in\mathcal{B}} \sum_{k\in\mathcal{B}, k\neq n} \mathbf{O}_n \odot \mathbf{O}_m \tag{12}$$

which is exactly the negative half value of the aggregate interference level of all the active users.

Now, suppose that an active player n unilaterally changes its channel sensing order from $\mathbf{O}_n$ to $\mathbf{O}_n^*$ while all other active players keep their sensing orders unchanged, then the change in player n's utility function is given by

$$\Delta U = U1_n\left(\mathcal{B}, \mathbf{O}_n^*, \mathbf{O}_{-n}\right) - U1_n(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n}). \tag{13}$$

To calculate the change in the potential function caused by the unilateral action change of player n, we define two player sets as follows:

$$\begin{aligned} \mathcal{U}_n &= (m \in \mathcal{B} : \mathbf{O}_n \odot \mathbf{O}_m > 0) \\ \mathcal{U}_n^* &= (m \in \mathcal{B} : \mathbf{O}_n^* \odot \mathbf{O}_m > 0). \end{aligned} \tag{14}$$

It is seen that $\mathcal{U}_n$ represents the set of active players that have overlapping sensing orders with player n when its action is $\mathbf{O}_n$, and $\mathcal{U}_n^*$ represents the set of players that have overlapping sensing orders with player n when its action is changed to $\mathbf{O}_n^*$. Based on the classification, all players except player n can be divided into the following exclusive four sets.

- $\mathcal{J}_1 = \mathcal{U}_n \cap (\mathcal{B} - \mathcal{U}_n^*)$: Active players in this set only have overlapping sensing orders with player n *before* it unilaterally changes its action.
- $\mathcal{J}_2 = (\mathcal{B} - \mathcal{U}_n) \cap \mathcal{U}_n^*$: Active players in this set only have overlapping sensing orders with player n *after* it unilaterally changes its action.
- $\mathcal{J}_3 = \mathcal{U}_n \cap \mathcal{U}_n^*$: Active players in this set have overlapping sensing orders with player n *both before and after* it unilaterally changes its action.
- $\mathcal{J}_4 = (\mathcal{B} - \mathcal{U}_n) \cap (\mathcal{B} - \mathcal{U}_n^*)$: Active players in this set have overlapping sensing orders with player n *neither before nor after* it unilaterally changes its action.

The change in the utility functions of players in the given four sets can be calculated as follows:

$$\Delta U_m = \begin{cases} \mathbf{O}_n \odot \mathbf{O}_m & \forall\, m \in \mathcal{J}_1 \\ -\mathbf{O}_n^* \odot \mathbf{O}_m & \forall\, m \in \mathcal{J}_2 \\ \mathbf{O}_n \odot \mathbf{O}_m - \mathbf{O}_n^* \odot \mathbf{O}_m & \forall\, m \in \mathcal{J}_3 \\ 0 & \forall\, m \in \mathcal{J}_4. \end{cases} \tag{15}$$

Based on the given user set classification, the change in the player n's utility function, as specified by (13), can be also expressed as follows:

$$\begin{aligned} \Delta U_n &= U_n\left(\mathcal{B}, \mathbf{O}_n^*, \mathbf{O}_{-n}\right) - U_n(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n}) \\ &= \sum_{k\in\mathcal{J}_1} \mathbf{O}_n \odot \mathbf{O}_k - \sum_{k\in\mathcal{J}_2} \mathbf{O}_n^* \odot \mathbf{O}_k \\ &\quad + \sum_{k\in\mathcal{J}_3} \left(\mathbf{O}_n \odot \mathbf{O}_k - \mathbf{O}_n^* \odot \mathbf{O}_k\right). \end{aligned} \tag{16}$$

Moreover, the change in the potential function caused by player n's unilateral action change is given by

$$\Delta\Phi = \Phi\left(\mathcal{B}, \mathbf{O}_n^*, \mathbf{O}_{-n}\right) - \Phi(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n})$$

$$= \frac{1}{2}\left(\Delta U_n + \sum_{m\in\mathcal{J}_1}\Delta U_m + \sum_{m\in\mathcal{J}_2}\Delta U_m + \sum_{m\in\mathcal{J}_3}\Delta U_m + \sum_{m\in\mathcal{J}_4}\Delta U_m\right). \quad (17)$$

Combining (15) and (17) yields the following equation:

$$\Delta\Phi = \sum_{m\in\mathcal{J}_1}\mathbf{O}_n\odot\mathbf{O}_m - \sum_{m\in\mathcal{J}_2}\mathbf{O}_n^*\odot\mathbf{O}_m + \sum_{m\in\mathcal{J}_3}\left(\mathbf{S}_n\odot\mathbf{S}_m - \mathbf{S}_n^*\odot\mathbf{S}_m\right). \quad (18)$$

Then, $\forall n\in\mathcal{B}$ and $\forall\,\mathbf{O}_n,\mathbf{O}_n^*\in\mathcal{A}_n$, the following equation always holds:

$$U1_n\left(\mathcal{B},\mathbf{O}_n^*,\mathbf{O}_{-n}\right) - U1_n(\mathcal{B},\mathbf{O}_n,\mathbf{O}_{-n}) = \Phi1\left(\mathcal{B},\mathbf{O}_n^*,\mathbf{O}_{-n}\right) - \Phi1(\mathcal{B},\mathbf{O}_n,\mathbf{O}_{-n}) \quad (19)$$

which can be directly obtained from (16) and (18). In other words, the change in the utility function caused by any player's unilateral action change is the same with that in the potential function. According to the definition given in [43], it is known that $\mathcal{G}_1$ is an exact potential game for every active-user set $\mathcal{B}$.

Based on the relationship between the aggregate interference level and the potential function, as specified by (5) and (12), respectively, it is seen that any optimal solution of problem $\mathbf{P1}$ is a global maximizer of the potential function. Furthermore, exact potential games have several promising attributes, and the most important two are as follows: 1) Every exact potential game has at least one pure strategy NE point, and 2) any global or local maximizer of the potential function constitutes a pure strategy NE of the game. Therefore, Theorem 1 follows. ■

Theorem 1 characterizes the general relationship between the original optimization problem $\mathbf{P}_1$ and the formulated state-based one-shot game $\mathcal{G}_1$, which is suitable for all scenarios with any active-user set. The following lemma further shows that there always exists an interference-free order profile for any active-user set $\mathcal{B}$.

Lemma 1: For any active-user set, i.e., $\forall\,\mathcal{B}\in\Gamma$, there exists a channel sensing order profile that is interference free.

Proof: We prove this lemma by construction. Without loss of generality, denote $\mathbf{B}_1=\{b_{11},b_{12},\ldots,b_{1M}\}$ as a channel sensing order vector that is arbitrarily chosen from the permutation set of M, i.e., $\mathbf{B}_1\in\mathcal{O}$. We construct $\mathbf{B}_2$ by a cyclic shift of $\mathbf{B}_1$, i.e., $\mathbf{B}_2=\{b_{12},b_{13},\ldots,b_{1M},b_{11}\}$. Iteratively, we construct $\mathbf{B}_k$ by a cyclic shift of $\mathbf{B}_{k-1}$, $k=3,4,\ldots,M$. We write the set of $\mathbf{B}_k$, $k=1,\ldots,M$, in a matrix form, i.e.,

$$\mathbf{B}_{\rm cs}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1(M-1)} & b_{1M}\\ b_{12} & b_{13} & \cdots & b_{1M} & b_{11}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ b_{1(M-1)} & b_{1M} & \cdots & b_{1(M-3)} & b_{1(M-2)}\\ b_{1M} & b_{11} & \cdots & b_{1(M-2)} & b_{1(M-1)}\end{bmatrix}. \quad (20)$$

$$\begin{bmatrix}1&2&3&4\\2&3&4&1\\3&4&1&2\\4&1&2&3\end{bmatrix}\quad\begin{bmatrix}1&3&2&4\\3&2&4&1\\2&4&1&3\\4&1&3&2\end{bmatrix}\quad\begin{bmatrix}1&4&2&3\\4&2&3&1\\2&3&1&4\\3&1&4&2\end{bmatrix}$$

$$\begin{bmatrix}1&2&4&3\\2&4&3&1\\4&3&1&2\\3&1&2&4\end{bmatrix}\quad\begin{bmatrix}1&3&4&2\\3&4&2&1\\4&2&1&3\\2&1&3&4\end{bmatrix}\quad\begin{bmatrix}1&4&3&2\\4&3&2&1\\3&2&1&4\\2&1&4&3\end{bmatrix}$$

Fig. 2. For $M=6$, there are six cyclic-shift matrices that correspond to interference-free order selection profiles.

For presentation, $\mathbf{B}_{\rm cs}$ is called the *cyclic-shift matrix*. It is noted that the following equation always holds:

$$\mathbf{B}_i\odot\mathbf{B}_j=0\qquad\forall\,i,j\in\{1,2,\ldots,M\},i\neq j \quad (21)$$

which further leads to the following equation:

$$\sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M}\mathbf{B}_i\odot\mathbf{B}_j=0. \quad (22)$$

Thus, $\{\mathbf{B}_1,\mathbf{B}_2,\ldots,\mathbf{B}_M\}$ constitutes an interference-free profile for the potential user set $\mathcal{N}$. For an arbitrary active-user set $\mathcal{B}$, the active users can choose $|\mathcal{B}|$ distinct order vectors among $\mathbf{B}_{\rm cs}$, which is also interference free. Therefore, Lemma 1 is proved. ■

It is noted that $\mathbf{B}_i$, $i=2,3,\ldots,M$, are associated with $\mathbf{B}_1$. Thus, the number of the given cyclic-shift matrices is equal to the number of different $\mathbf{B}_1$. Considering that all the channel sensing orders in an interference-free profile can be obtained by a cyclic shift of each other, we can generate different $\mathbf{B}_1$ by fixing one element while permuting all other elements. Thus, the total number is given by the permutation number of $M-1$, i.e., $(M-1)!$. For example, there are six cyclic-shift matrices of such interference-free profiles for $M=4$, which are shown in Fig. 2.

Based on Lemma 1, we can study the achievable performance of the formulated state-based one-shot game.

Proposition 1: For any active-user set, i.e., $\forall\,\mathcal{B}\in\Gamma$, the best pure strategy NE of $\mathcal{G}_1$ corresponds to an interference-free sensing order profile.

Proof: For $\forall\,\mathcal{B}\in\Gamma$, there always exists a channel sensing order profile that is interference free. For example, one can allocate different rows of a cyclic-shift matrix, as specified by (20), to the active users. Clearly, an interference-free channel sensing order profile is optimal to problem $\mathbf{P1}$. Thus, according to Theorem 1, Proposition 1 follows. ■

The given theoretical analysis characterizes the underlying relationships between the original optimization problem $\mathbf{P}1$ and the formulated state-based one-shot order selection game $\mathcal{G}_1$. The results are promising since the best pure strategy NE of $\mathcal{G}_1$ corresponds to an optimal solution of $\mathbf{P}1$, as shown by Proposition 1. However, the state-based one-shot games cannot be solved as the current active-user sets are random and unknown to the players. More specifically, each player only knows

its state (active or inactive) but does not know whether the other players are active or not.

Note that the given game model is similar to the traditional coloring game, in which the players selecting the same color conflict with each other. However, there are two key differences: 1) In a coloring game, the interactive relationship is binary, i.e., the color of a player is the same with that of another player or they are different; in contrast, the sensing orders in the formulated game may fully or partially overlap, which requires new formulation and analysis; 2) the set of active users in our work is time varying, whereas that in traditional coloring games is fixed.

B. Robust Order Selection Game

Here, we formulate the robust order selection game, which can be regarded as the expectation over all system states of the given one-shot order selection game [23]. Formally, the robust order selection game is denoted as $\mathcal{G}_2 = \{\mathcal{N}, \{\mathcal{A}_n\}_{n\in\mathcal{N}}, \{U2_n\}_{n\in\mathcal{N}}\}$, where $\mathcal{N}$ is the player set, $\mathcal{A}_n$ is the action set, and $U2_n$ is the utility function of player n. Note that the players in $\mathcal{G}_2$ are all potential users. The utility function is defined as follows:

$$\begin{aligned} U2_n(\mathbf{O}_n, \mathbf{O}_{-n}) &= \mathrm{E}_{\mathcal{B}}\left[U1_n(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n})\right] \\ &= \sum_{\mathcal{B}\in\Gamma} \mu(\mathcal{B}) U1_n(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n}) \end{aligned} \quad (23)$$

where $\mathrm{E}_{\mathcal{B}}$ is the operation of taking expectation over all system states, and $U1_n(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n})$ is the utility function of the state-based one-shot game $\mathcal{G}_1$, as characterized by (8). Similarly, the robust channel sensing order selection game can be expressed as

$$\mathcal{G}_2: \quad \max\ U2_n(\mathbf{O}_n, \mathbf{O}_{-n}) \qquad \forall\, n \in \mathcal{N}. \quad (24)$$

The properties of the NE of the formulated robust order selection game is characterized by the following theorem.

Theorem 2: The robust order selection game $\mathcal{G}_2$ is also an exact potential game that has at least one pure strategy NE point. More importantly, any optimal solution of problem $\mathbf{P2}$ constitutes a pure strategy NE of $\mathcal{G}_2$.

Proof: We define the potential function as follows:

$$\Phi 2(\mathbf{O}_n, \mathbf{O}_{-n}) = \mathrm{E}_{\mathcal{B}}\left[\Phi 1(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n})\right] \quad (25)$$

where $\Phi 1(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n})$ is the potential function of the state-based one-shot game $\mathcal{G}_1$. According to (12) and (25), it is noted that $\Phi 2(\mathbf{O}_n, \mathbf{O}_{-n})$ is exactly the negative value of the expected aggregate interference level of all the potential users.

Now, suppose that an arbitrary player n unilaterally changes its channel sensing order from $\mathbf{O}_n$ to $\mathbf{O}_n^*$ while all other active players keep their sensing orders unchanged, then the change in player n's utility function is given by

$$\begin{aligned} &U2_n\left(\mathbf{O}_n^*, \mathbf{O}_{-n}\right) - U2_n(\mathbf{O}_n, \mathbf{O}_{-n}) \\ &= \sum_{\mathcal{B}\in\Gamma} \mu(\mathcal{B}) \left[U1_n\left(\mathcal{B}, \mathbf{O}_n^*, \mathbf{O}_{-n}\right) - U1_n(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n})\right]. \end{aligned} \quad (26)$$

Accordingly, the change in the potential function by the action change of player n is given by

$$\begin{aligned} &\Phi 2_n\left(\mathbf{O}_n^*, \mathbf{O}_{-n}\right) - \Phi 2_n(\mathbf{O}_n, \mathbf{O}_{-n}) \\ &= \sum_{\mathcal{B}\in\Gamma} \mu(\mathcal{B}) \left[\Phi 1\left(\mathcal{B}, \mathbf{O}_n^*, \mathbf{O}_{-n}\right) - \Phi 1(\mathcal{B}, \mathbf{O}_n, \mathbf{O}_{-n})\right]. \end{aligned} \quad (27)$$

Now, applying the result obtained in (19) into (26) and (27) yields the following equations:

$$\begin{aligned} &U2_n\left(\mathbf{O}_n^*, \mathbf{O}_{-n}\right) - U2_n(\mathbf{O}_n, \mathbf{O}_{-n}) \\ &\quad = \Phi 2\left(\mathbf{O}_n^*, \mathbf{O}_{-n}\right) - \Phi 2(\mathbf{O}_n, \mathbf{O}_{-n}) \qquad \forall\, n \in \mathcal{N}. \end{aligned} \quad (28)$$

Therefore, the robust order selection game is also an exact potential game that admits at least one pure strategy NE. In addition, it is seen that the optimal solution of $\mathbf{P2}$ is a global maximizer of $\mathcal{G}_2$. Therefore, Theorem 2 is proved. ■

Similarly, the following proposition characterizes the achievable performance of the robust order selection game.

Proposition 2: The best pure strategy NE of $\mathcal{G}_2$ corresponds to an interference-free sensing order profile.

Proof: Based on Lemma 1, similar lines for the proof of Proposition 1 can be applied to prove this statement. ■

The relationship between the robust order selection game $\mathcal{G}_2$ and the state-based one-shot game $\mathcal{G}_1$ is characterized by the following theorem.

Theorem 3: A best pure strategy NE of the robust order selection game $\mathcal{G}_2$ is also a best pure strategy NE of the state-based one-shot game $\mathcal{G}_1$ for all system states, i.e., $\forall\, \mathcal{B} \in \Gamma$.

Proof: According to the given analysis, it is known that any best pure strategy NE of $\mathcal{G}_2$ is interference free, which means that the channel sensing order vectors of any two potential users are orthogonal. For any active-user set $\mathcal{B}$ in the state-based game $\mathcal{G}_1$, suppose that the active users employ actions that are drawn from a best pure strategy NE of $\mathcal{G}_2$. As a result, the channel sensing order vector profile of the active users is also orthogonal. According to Proposition 1, the channel sensing orders correspond to a pure strategy NE of $\mathcal{G}_1$ with arbitrary active-user set $\mathcal{B}$. Therefore, Theorem 3 is proved. ■

As previously analyzed, it seems that finding the best pure strategy NE of $\mathcal{G}_1$ using traditional approaches is impossible since the inherent system state $\mathcal{B}$ is unknown and changing from slot to slot. Moreover, finding the best pure strategy NE of $\mathcal{G}_2$ seems impossible since the distribution probabilities of active-user sets are unknown, and there is no information exchange between users. In the following, we propose a distributed learning approach that asymptotically converges to the best pure strategy NE of $\mathcal{G}_2$, which also converges to the best strategy NE of $\mathcal{G}_1$ according to Theorem 3.

V. Distributed Learning Approach With Random Active-User Set

Generally, potential games enjoy good convergence properties. Specifically, there are some commonly used learning algorithms in the literature, which converge to pure strategy NE points of potential games, e.g., best (better) response [43], spatial adaptive play [28], log-linear learning [44], [45], and no-regret learning [25], [27]. However, these algorithms cannot

be applied in the considered dynamic network since they are originally designed for static game models and need information exchange among the players. Recently, two learning algorithms for the problems of channel selection in CR networks with dynamic spectrum opportunities [46] and in canonical networks with block-fading channels [32] were proposed. Although the learning algorithms therein considered the dynamic environment, they are for game models with a fixed number of players and, hence, cannot be applied to the considered dynamic CR networks with changing number of active players.

In this paper, we propose a stochastic learning algorithm for game models with changing number of active-user sets in each slot. For presentation, the slot index is added into the game models: $s_n(k)$ represents the state of player n in slot k, $\mathcal{B}(k)$ is the active-user set in slot k, and $r_n(k)$ is the received binary feedback of player n in slot k. According to the transmission strategies of the users, the binary feedback is determined as follows: 1) $r_n(k)=1$: the transmission of player n is successful; and 2) $r_n(k)=0$: player n experiences a collision or does find idle channels. We assume that there is a coordination mechanism between the players such that they use a common cyclic-shift matrix as their action space,[4] i.e., $\mathcal{A}_n=\mathbf{B}_{\rm cs},\ \forall\, n\in\mathcal{N}$.

To begin with, the players employ mixed strategies in each slot. Specifically, $\mathbf{Q}(k)=(\mathbf{q}_1(k),\ldots,\mathbf{q}_n(k))$ denotes the mixed-strategy profile in slot k, in which $\mathbf{q}_n(k)=(q_{n1}(k),\ldots,q_{nM}(k))$ is the probability vector of player n choosing each channel sensing order. The underlying ideas of the proposed stochastic learning algorithm can be summarized as follows: 1) In the first slot, player n being active, it chooses the channel sensing orders with equal probabilities $\mathbf{q}_n(k_0)=((1/M),\ldots,(1/M))$; 2) for an active player $n\in\mathcal{B}(k)$ in slot k, it receives binary feedback $r_n(k)$ at the end of the slot and employs a rule to update its mixed strategy. For an inactive user, it gets zero and keeps its mixed strategy unchanged. The proposed stochastic learning algorithm is formally described in Algorithm 1.

Algorithm 1: Stochastic learning algorithm with randomly changing active-user set

---

Loop for $k=0,1,2,\ldots$,

1. **Selecting channel sensing orders stochastically**: At the beginning of slot k, each active player $n\in\mathcal{B}(k)$ selects a channel sensing order $a_n(k)\in\mathcal{A}_n$ according to its mixed strategy $\mathbf{q}_n(k)$.
2. **Accessing and receiving binary feedback**: All the active players perform sequential channel sensing and access. Specifically, they sense the licensed channels one by one according to its chosen order and transmit data in the first idle channel. At the end of slot k, each player receives a binary feedback $r_n(k)$, which is jointly determined by the activities of PUs and the channel sensing orders of other active users.
3. **Updating mixed strategy**: All the active users update their mixed strategies according to the following rule:

$$\mathbf{q}_n(k+1)=\mathbf{q}_n(k)+br_n(k)\left(\mathbf{I}_{a_n(k)}-\mathbf{q}_n(k)\right) \tag{29}$$

where $0<b<1$ is the learning parameter, $\mathbf{I}_{a_n(k)}$ is a unit vector with the $a_n(k)$th component being one. The inactive users keep their mixed strategies unchanged.

End loop

---

[4]The reason for making this assumption is that the action space is huge, e.g., the size of the action space for $M=5,6$ are 120 and 720, respectively. Thus, an approach for reducing the action space size is needed to accelerate the convergence speed. However, it should be pointed that the proposed learning stochastic algorithm is also suitable for the nonreducing action space.

The convergence of the proposed stochastic learning in the presence of a changing active-user set is characterized by the following theorem.

Theorem 4: When all the players use a common cyclic-shift matrix as the action space, the proposed stochastic learning algorithm asymptotically converges to the best pure strategy NE points of both $\mathcal{G}_1$ and $\mathcal{G}_2$ if the learning parameter goes sufficiently small, i.e., $b\to 0$.

Proof: The proof of Theorem 4 is organized as follows. First, using the results of replicator dynamic [49], we show that the long-term behavior of the sequence $\mathbf{Q}(k)$ can be approximately characterized by an ordinary differential equation (ODE). Second, by investigating the convergence of the potential function with regard to the mixed-strategy profile, i.e., $\Phi(\mathbf{Q}(k))$, it is shown that the proposed stochastic learning algorithm asymptotically converges to the Nash strategy of the robust game $\mathcal{G}_2$. Thus, we have (30)–(33), shown at the bottom of the next page.

1) Associated ODE: It is seen that the update rule shown in (29) is for active players. We first extend the update rule for all the potential players as follows:

$$\mathbf{q}_n(k+1)=\mathbf{q}_n(k)+br_n(k)s_n(k)\left(\mathbf{I}_{a_n(k)}-\mathbf{q}_n(k)\right) \tag{34}$$

where $s_n(k)$ indicates the event that player n is active $(s_n(k)=1)$ or inactive $(s_n(k)=0)$ in slot k. To investigate the evolution of the mixed-strategy profile of all potential players, we rewrite the given update rule as follows:

$$\mathbf{Q}(k+1)=\mathbf{Q}(k)+bG\left(\mathbf{Q}(k),a(k),r(k),s(k)\right) \tag{35}$$

where $G(\cdot,\cdot,\cdot,\cdot)$ is the update rule specified by (34). For presentation, we denote the conditional expected value of function G as follows:

$$F(\mathbf{Q})=\mathrm{E}\left[G\left(\mathbf{Q}(k),a(k),r(k),s(k)\right)|\mathbf{Q}(k)\right]. \tag{36}$$

Following the idea of stochastic approximation [47], the long-term behavior of the mixed-strategy profiles $\mathbf{Q}(k)$ is characterized by the following lemmas.

Lemma 2: When the learning parameter goes sufficiently small, i.e., $b\to 0$, the sequence of the mixed-strategy profile $\mathbf{Q}(k)$ converges to the solution of the following ODE with initial value $\mathbf{Q}(0)$:

$$\frac{d\mathbf{Q}}{dt}=F(\mathbf{Q}). \tag{37}$$

Proof: By the method of interpolation, similar lines given in [48] can be applied to prove this lemma. ■

Lemma 3: If the learning parameter b is sufficiently small, the following are true for the proposed stochastic learning algorithm.

1) All Nash equilibria of the game are stationary points of (37).
2) All stationary points of (37) that are not stable are not Nash equilibria.

Proof: See [48]. ■

According to the given analysis, we can conclude that the proposed stochastic learning algorithm would converge to Nash equilibria if the sequence $\mathbf{Q}(k)$ converges to the stationary points of (37). In the following, we analyze the asymptotical convergence behavior of the proposed stochastic learning algorithm.

2) Asymptotical Convergence Behavior: For presentation, we denote the active-player set excluding player n as $\mathcal{B}_n = \{k \in \{\mathcal{N} \setminus n\} : s_k = 1\}$ and all the possible profiles of $\mathcal{B}_n$ as Γ_n. We define $h_n(m, \mathbf{Q}_{-n})$ as the expected payoff of player n if it chooses a pure strategy m, i.e., $a_n = \mathbf{B}_m$, while all other active players employ mixed strategies $\mathbf{Q}_{-n}$. Mathematically, the analytical expression of $h_n(m, \mathbf{Q}_{-n})$ is given in (30). Similarly, the expected value of the potential function $H(\mathbf{Q})$ over mixed-strategy profile $\mathbf{Q}$ and the expected value $H_n(m, \mathbf{Q}_{-n})$ when player n chooses a pure strategy m while all other active players employ mixed strategies $\mathbf{Q}_{-n}$ is given by (31) and (32), respectively. Since $H(\mathbf{Q}) = \sum_m q_{nm} H_n(m, \mathbf{Q}_{-n})$, the variation of $H(\mathbf{Q})$ can be expressed as follows:

$$\frac{\partial H(\mathbf{Q})}{\partial q_{nm}} = H_n(m, \mathbf{Q}_{-n}). \tag{38}$$

We can rewrite the ODE specified by (37) as follows:

$$\frac{dq_{nm}}{dt} = F_{nm}(\mathbf{Q}) \qquad \forall\, n \in \mathcal{N}. \tag{39}$$

According to (39) and (30), the given equation can be further rewritten as follows:

$$\begin{aligned}\frac{dq_{nm}}{dt} &= \lambda_n \Bigg(q_{nm}(1 - q_{nm}) \mathrm{E}_{\mathcal{B}_n, \mathbf{Q}_{-n}} [r_n | (\mathcal{B}_n, m, \mathbf{Q}_{-n})] \\ &\qquad + \sum_{k \neq m} q_{nk}(-q_{nm}) \mathrm{E}_{\mathcal{B}_n, \mathbf{Q}_{-n}} [r_n | (\mathcal{B}_n, k, \mathbf{Q}_{-n})] \Bigg) \\ &= \lambda_n q_{nm} \left(h_n(m, \mathbf{Q}_{-n}) - \sum_k q_{nk} h_n(k, \mathbf{Q}_{-n}) \right) \\ &= \lambda_n q_{nm} \sum_k q_{nk} \left(h_n(m, \mathbf{Q}_{-n}) - h_n(k, \mathbf{Q}_{-n}) \right). \end{aligned} \tag{40}$$

Based on the given analysis, we study the long-term behavior of the potential function $H(\mathbf{Q})$. Specifically, the derivative of $H(\mathbf{Q})$ is given by (33). According to the properties of the potential games, the following equation always holds:

$$\begin{aligned} &H_n(m, \mathbf{Q}_{-n}) - H_n(k, \mathbf{Q}_{-n}) \\ &\qquad = h_n(m, \mathbf{Q}_{-n}) - h_n(k, \mathbf{Q}_{-n}) \qquad \forall\, n, m, k. \end{aligned} \tag{41}$$

Therefore, (33) can be further expressed as follows:

$$\begin{aligned} \frac{dH(\mathbf{Q})}{dt} &= \frac{1}{2} \sum_{n,m,k} \lambda_n q_{nm} q_{nk} \\ &\qquad \times \left(h_n(m, \mathbf{Q}_{-n}) - h_n(k, \mathbf{Q}_{-n}) \right)^2 \geq 0. \end{aligned} \tag{42}$$

The given equation shows that $H(\mathbf{Q})$ increases as the algorithm iterates. Furthermore, it is known that $H(\mathbf{Q})$ is bounded by $H(\mathbf{Q}) \leq 0$. Therefore, $H(\mathbf{Q})$ will eventually converge to

---

$$h_n(m, \mathbf{Q}_{-n}) = \mathrm{E}_{\mathcal{B}_n, \mathbf{Q}_{-n}} [r_n | (\mathcal{B}_n, m, \mathbf{Q}_{-n})] = \sum_{\mathcal{B}_n \in \Gamma_n} \prod_{k \in \mathcal{B}_n} \lambda_k \left(\sum_{a_k \in \mathcal{A}_k, k \in \mathcal{B}_n} r_n(a_1, \ldots, a_{n-1}, \mathbf{B}_m, a_{n+1}, \ldots, a_N) \prod_{k \in \mathcal{B}_n} q_{k a_k} \right) \tag{30}$$

$$H(\mathbf{Q}) = \mathrm{E}_{\mathcal{B}_n, \mathbf{Q}} [\Phi | (\mathcal{B}_n, \mathbf{q}_n, \mathbf{Q}_{-n})] = \sum_{\mathcal{B} \in \Gamma} \prod_{k \in \mathcal{B}} \lambda_k \left(\sum_{a_k \in \mathcal{A}_k, k \in \mathcal{B}_n} \Phi(a_1, \ldots, a_{n-1}, a_n, a_{n+1}, \ldots, a_N) \prod_{k \in \mathcal{B}} q_{k a_k} \right) \tag{31}$$

$$H_n(m, \mathbf{Q}_{-n}) = \mathrm{E}_{\mathcal{B}_n, \mathbf{Q}_{-n}} [\Phi | (\mathcal{B}_n, m, \mathbf{Q}_{-n})] = \sum_{\mathcal{B}_n \in \Gamma_n} \prod_{k \in \mathcal{B}_n} \lambda_k \left(\sum_{a_k \in \mathcal{A}_k, k \in \mathcal{B}_n} \Phi(a_1, \ldots, a_{n-1}, \mathbf{B}_m, a_{n+1}, \ldots, a_N) \prod_{k \in \mathcal{B}_n} q_{k a_k} \right) \tag{32}$$

$$\begin{aligned} \frac{dH(\mathbf{Q})}{dt} &= \sum_{n,m} \frac{\partial H(\mathbf{Q})}{q_{nm}} \frac{dq_{nm}}{dt} = \sum_{n,m} H_n(m, \mathbf{Q}_{-n}) \lambda_n q_{nm} \sum_k q_{nk} \left(h_n(m, \mathbf{Q}_{-n}) - h_n(k, \mathbf{Q}_{-n}) \right) \\ &= \sum_{n,m,k} \lambda_n q_{nm} q_{nk} H_n(m, \mathbf{Q}_{-n}) \left(h_n(m, \mathbf{Q}_{-n}) - h_n(k, \mathbf{Q}_{-n}) \right) \\ &= \sum_{n,m,k} \lambda_n q_{nm} q_{nk} H_n(k, \mathbf{Q}_{-n}) \left(h_n(k, \mathbf{Q}_{-n}) - h_n(m, \mathbf{Q}_{-n}) \right) \\ &= \frac{1}{2} \sum_{n,m,k} \lambda_n q_{nm} q_{nk} \left(H_n(m, \mathbf{Q}_{-n}) - H_n(k, \mathbf{Q}_{-n}) \right) \left(h_n(m, \mathbf{Q}_{-n}) - h_n(k, \mathbf{Q}_{-n}) \right) \end{aligned} \tag{33}$$

some maximum points, when $(dH(\mathbf{Q}))/dt = 0$. Finally, we have the following relationships:

$$\begin{aligned}
&\frac{dH(\mathbf{Q})}{dt} = 0 \\
&\Rightarrow h_n(m, \mathbf{Q}_{-n}) - h_n(k, \mathbf{Q}_{-n}) = 0 \qquad \forall\, n, m, k \\
&\Rightarrow \frac{dq_{nm}}{dt} = 0 \qquad \forall\, n, m \\
&\Rightarrow \frac{d\mathbf{Q}}{dt} = 0 \\
&\Rightarrow \mathbf{Q} \text{ converges to the stationary point of (37).} \qquad (43)
\end{aligned}$$

Therefore, according to Lemmas 2 and 3, it is proved that the proposed stochastic learning algorithm converges to Nash equilibria of the games. Moreover, it is noted that all the players use a common cyclic-shift matrix as the action set; therefore, according to Propositions 1 and 2 and Theorem 3, we conclude that the proposed stochastic learning algorithm converges to the best NE of $\mathcal{G}_1$ and $\mathcal{G}_2$. Therefore, Theorem 4 is proved. ■

The proposed stochastic learning algorithm is promising since it asymptotically converges to the best NE of $\mathcal{G}_1$ and $\mathcal{G}_2$. It is fully distributed and autonomous. More importantly, it captures the changing number of active players and the dynamic spectrum opportunities well. Based on the given analysis, it is seen that it also achieves the best solutions of the original problems $\mathbf{P}1$ and $\mathbf{P}2$, respectively.

Remark 2: It is emphasized that the proposed learning stochastic algorithm is online. Specifically, the active users perform sequential channel sensing and access, receive binary feedbacks, and then adjust their mixed strategies. Therefore, they transmit data before and after the learning algorithm converges.

Remark 3: The complexity of the proposed learning algorithm is very low. In particular, an active user only needs to record its current mixed strategy, the current chosen order, and the received payoffs; furthermore, the update rule is linear. In addition, the inactive users keep their mixed strategies unchanged.

VI. Simulation Results and Discussion

In the simulation study, we denote the sensing time fraction in a slot as $\tau = T_s/T$. As a result, the normalized achievable throughput of an active player, which successfully accesses an idle channel in a slot, is given by $R = 1 - n\tau$. Following the similar setting in [40], the slot length is set to $T = 100$ ms, and the sensing duration is set to $T_s = 5$ ms. It is assumed that energy detection is employed by each SU, then the spectrum sensing performance is characterized by $P_d(T_s) = 0.9$ and $P_f(T_s) = 0.1$, which is determined by the detection threshold [40]. Furthermore, for convenience of discussion, we assume that the idle probabilities of all licensed channels are the same, i.e., $\theta_m = \theta$, $\forall\, m \in \{1, 2, \ldots, M\}$, and the active probabilities of each user are also the same, i.e., $\lambda_n = \lambda$, $\forall\, n \in \{1, 2, \ldots, N\}$.

A. Convergence Behavior

Here, we study the convergence behaviors of the proposed stochastic learning algorithm in the presence of dynamic active users. For presentation, it is assumed that there are five licensed channels and five potential users, i.e., $M = 5$ and $N = 5$. The licensed channel idle probabilities are set to $\theta = 0.6$, and the user active probabilities are set to $\lambda = 0.5$. The learning parameter is set to $b = 0.05$, which has been optimized by experiment.

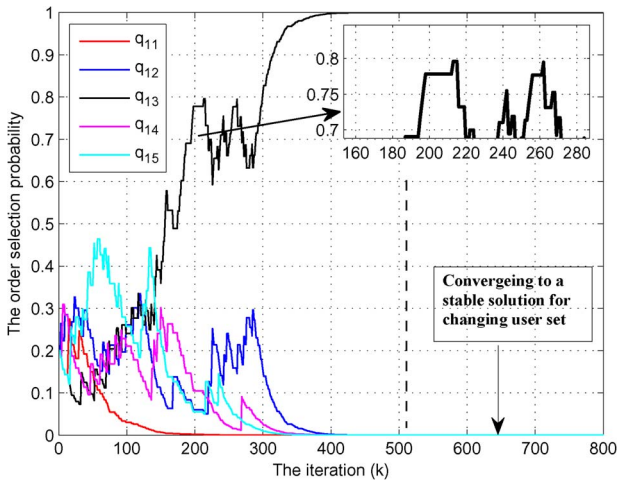

Fig. 3. Evolution of the order selection probabilities of arbitrarily chosen users ($M = 5$, $N = 5$, $\theta = 0.6$, $\lambda = 0.5$).

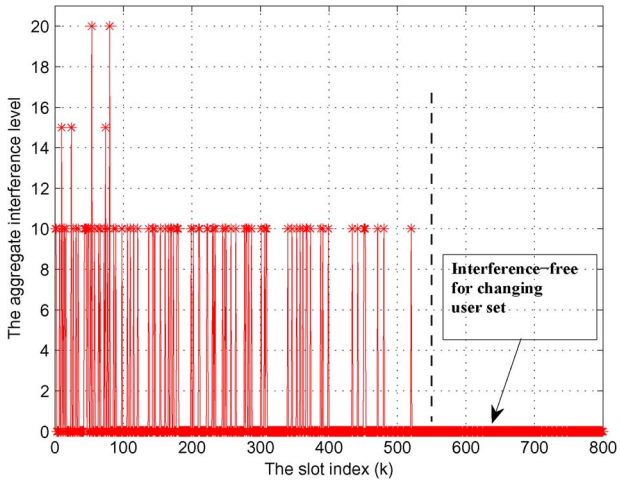

Fig. 4. Evolution of the aggregate interference level (the number of license channels is $M = 5$, the number of SUs is $N = 5$, the idle probabilities of the licensed channel are set to $\theta_n = 0.6$, and the active probabilities of the SUs in each slot are set to $\lambda = 0.5$).

For an arbitrarily chosen user, the evolution of order selection probabilities is shown in Fig. 3. It is noted in the figure that the order selection probabilities remain unchanged in successive multiple slots (for example, from slot 200 to slot 215), which corresponds to the event that the player is inactive in these slots. It is seen that it finally converges to a pure strategy ($\mathbf{q} = \{0, 0, 1, 0, 0\}$) in about 400 iterations (slots). After slot 400, although the active-user set is randomly changing, the player employs the converging stable solution when it is active. The results shown in the figure validate the convergence of the proposed learning algorithm in the presence of the changing active-user set.

The evolution of the aggregate interference level is shown in Fig. 4. It is noted that the aggregate collision level finally

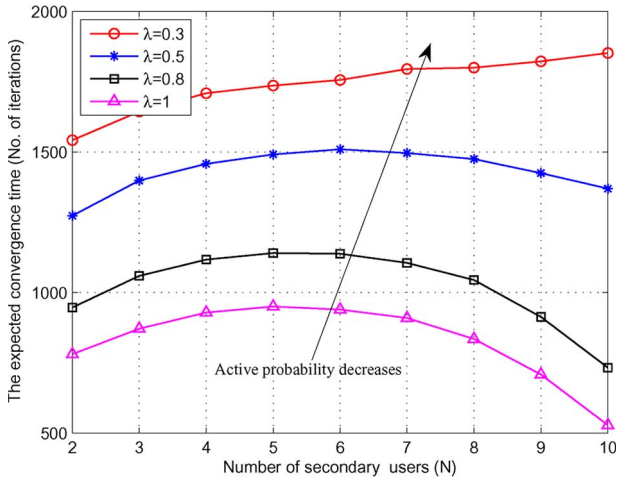

Fig. 5. Expected convergence time (No. of iterations) with different numbers of SUs and different active probabilities of the SUs in each slot (the number of licensed channels is $M = 10$, and the channel idle probabilities are set to $\theta_n = 0.8$).

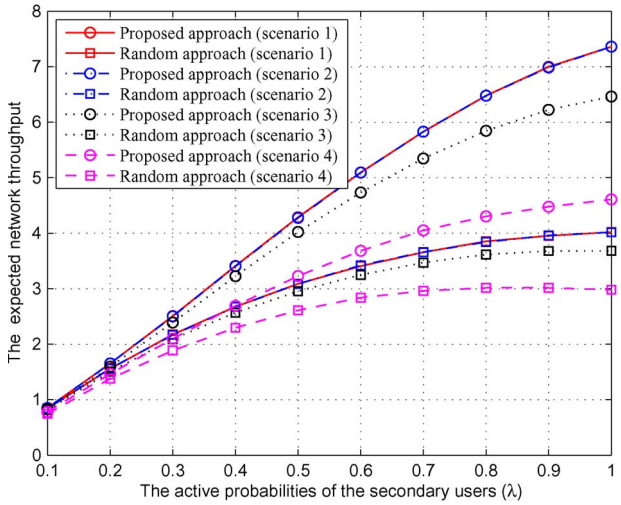

Fig. 6. Comparison results versus the active probabilities of the SUs in each slot (the number of licensed channels is $M = 10$, and the number of SUs is $N = 10$).

decreases to zero in about 500 iterations, which implies the convergence behaviors of all the potential users. More importantly, it is noted that the converging order selection profile is interference free for every changing active-user set.

It is seen that the proposed learning approach takes several slots to converge to stable solutions. Therefore, it is interesting to study its convergence time. We define the expected convergence time as the number of iterations where there exists a component of the mixed strategy of each player, which is sufficiently approaching one, e.g., larger than 0.95. The expected convergence time for a dynamic system consisting of ten licensed channels with different numbers of SUs and different active probabilities are shown in Fig. 5. It is noted in the figure that for a given number of SUs, e.g., $N = 5$, the expected convergence time increases as the user active probability decreases. The reason is that an SU with lower active probability performs learning occasionally, whereas an SU with higher active probability performs learning more frequently. Moreover, for scenarios with large active probabilities, e.g., $\lambda = 0.5, 0.8, 1$, the expected convergence time increases when the number of SUs is small, e.g., $N < 6$, and decreases when the number of SUs becomes large, e.g., $N \geq 6$. The reasons are as follows: 1) When the number of SUs is small, the resources are relatively abundant, and the users spend more time in finding the desirable sensing orders; and 2) when the number of SUs is large, the collision frequency becomes large.

B. Throughput Performance

Here, we evaluate the throughput performance of the proposed learning approach in different scenarios. We compare the throughput performance of the proposed learning approach with the random selection approach. In the random selection approach, each active player chooses the sensing and access order randomly and autonomously. In the considered distributed CRs with changing active-user set and time-varying spectrum opportunities, the random selection approach is an intuitive approach.

The simulated systems are with ten licensed channels. To study the effect of PU dynamics on the achievable system throughput, we consider the following four scenarios with different licensed idle probabilities.

1) (Scenario 1) Homogeneous licensed channels: All the idle probabilities are set to 0.8.
2) (Scenario 2) Slight heterogeneity of the licensed channels: The licensed channel idle probabilities are set to $\{0.7, 0.7, 0.7, 0.8, 0.8, 0.8, 0.8, 0.9, 0.9, 0.9\}$.
3) (Scenario 3) Moderate heterogeneity of the licensed channels: The licensed channel idle probabilities are set to $\{0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.8, 0.8, 0.9, 0.9\}$.
4) (Scenario 4) Heavy heterogeneity of the licensed channels: The licensed channel idle probabilities are set to $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.8, 0.9\}$.

The comparison results versus the active probabilities of the SUs are shown in Fig. 6. There are ten SUs, and all of them have the same active probability in each slot. The results are obtained by simulating 100 000 successive slots and then taking the expected value. It is seen that the normalized expected system throughput of both approaches increases as the active probabilities of SUs increase. In addition, the proposed learning approach outperforms the random selection approach for four considered scenarios. In particular, as the active probabilities increase, the throughput gap becomes significant. These results validate the proposed learning approach both for homogeneous and heterogeneous scenarios.

The comparison results versus the number of SUs are shown in Fig. 7. The active probability of all SUs in each slot is set to $\lambda = 0.7$. The results are obtained by simulating 100 000 successive slots and then taking the expected value. It is seen that the normalized expected system throughput of both approaches increases as the number of potential users increases. In addition, the proposed learning approach outperforms the random selection approach for four considered scenarios. In particular, as the number of potential users increases, the throughput gap becomes significant. These results again validate the proposed

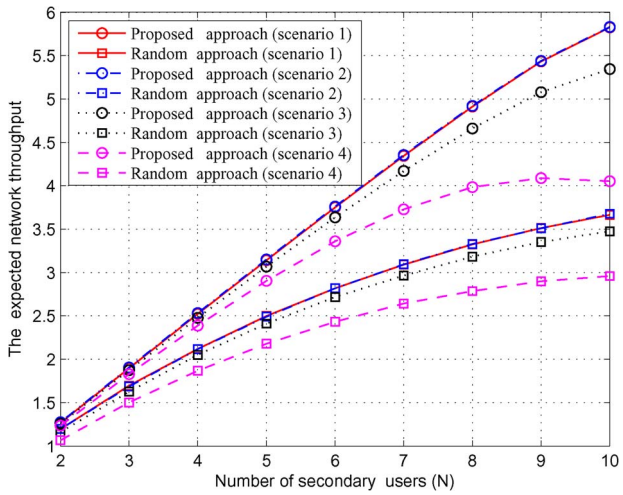

Fig. 7. Comparison results versus the number of SUs (the number of licensed channels is $M = 10$, and the active probability of each SU in each slot is $\lambda = 0.7$).

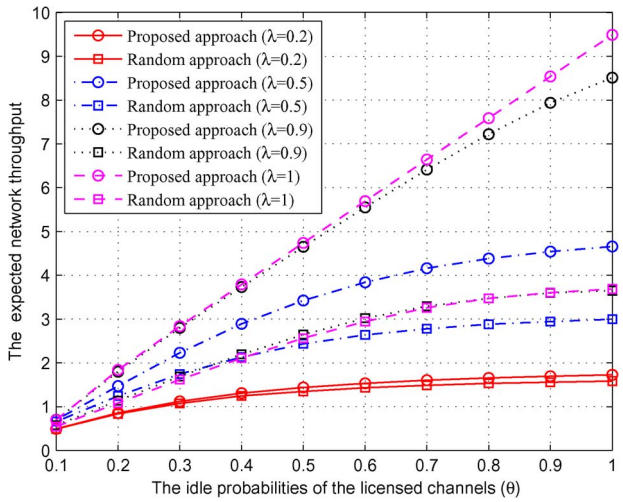

Fig. 8. Comparison results versus the idle probabilities of licensed channels ($M = 5$, $N = 5$, $\lambda = 0.8$).

learning approach both for homogeneous and heterogeneous scenarios.

The comparison results versus the idle probabilities of licensed channels are shown in Fig. 8. There are ten SUs; furthermore, all the idle probabilities of the licensed channels are set the same. The results are obtained by simulating 100 000 successive slots and then taking the expected value. It is seen that the normalized expected system throughput of both approaches increases as the idle probabilities of licensed channels increase. When the active probability of the SUs is small, e.g., $\lambda = 0.2$, the difference between the throughput performance of the proposed learning approach and that of the random approach is trivial. However, for larger active probabilities, e.g., $\lambda = 0.5, 0.8, 1$, the throughput gap is significant. Moreover, for a given active probability of the SUs, the throughput gap becomes significant as the idle probabilities of licensed channels increase.

To summarize, the simulation results validate the convergence of the proposed learning approach in dynamic CR networks with the changing set of active SUs and unknown system parameters. Moreover, it is shown that it achieves satisfactory throughput performance in both homogeneous and heterogeneous environments.

VII. Conclusion

We have studied the problem of multiuser sequential channel sensing and access in dynamic CR networks, in which the active-user set is randomly changing from slot to slot. Furthermore, each user only has its individual information, and information exchange among users is not available. The goal of the users is to determine channel order for sensing and access. We defined a generalized interference metric to address the overlapping of multiple channel orders and established two optimization objectives: minimizing the aggregate interference for each active-user set and minimizing the expected aggregate interference for all potential users. We proposed a state-based one-shot game and a robust game to solve the optimization problems. We proved that the best NE of the two games corresponds to the optimal solutions of the two optimization problems, respectively. To cope with the UDI information constraints in the distributed and dynamic networks, we proposed a stochastic learning algorithm, which was analytically proved to converge to Nash equilibria of the two formulated games in the presence of a changing active-player set. The convergence and superior performance of the proposed learning algorithm were validated by simulation results. In the future, we plan to consider dynamic CR networks with user mobility.

References

[1] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] C. Chou, N. S. Shankar, H. Kim, and K. G. Shin, "What and how much to gain by spectrum agility?" *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 576–588, Apr. 2007.

[4] S. Srinivasa and S. Jafar, "How much spectrum sharing is optimal in cognitive radio networks?" *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 4010–4018, Oct. 2008.

[5] Y. Zhao, S. Mao, J. Neel, and J. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, methodology," *Proc. IEEE*, vol. 97, no. 4, pp. 642–659, Apr. 2009.

[6] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.

[7] S. Hu, Y.-D. Yao, and Z. Yang, "MAC protocol identification using support vector machines for cognitive radio networks," *IEEE Wireless Commun.*, vol. 21, no. 1, pp. 52–60, Feb. 2014.

[8] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[9] Y. Xu *et al.*, "Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1689–1713, 2013.

[10] Z. Khan, J. Lehtomäki, L. DaSilva, and M. Latva-aho, "Autonomous sensing order selection strategies exploiting channel access information," *IEEE Trans. Mobile Comput.*, vol. 12, no. 2, pp. 274–288, Feb. 2013.

[11] J. Zhao and X. Wang, "Channel sensing order in multi-user cognitive radio networks," in *Proc. IEEE DySPAN*, 2012, pp. 397–407.

[12] A. Mendes, C. Augusto, M. Silva, R. M. Guedes, and J. F. de Rezende, "Channel sensing order for cognitive radio networks using reinforcement learning," in *Proc. 36th Annu. IEEE Conf. Local Comput. Netw.*, 2011, pp. 546–553.

[13] H. Shokri-Ghadikolaei, F. Sheikholeslami, and M. Nasiri-Kenari, "Distributed multiuser sequential channel sensing schemes in multichannel cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2055–2067, May 2013.
[14] M. Masonta, M. Mzyece, and N. Ntlatlapa, "Spectrum decision in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1088–1107, 2013.
[15] S. Kim and G. Giannakis, "Sequential and cooperative sensing for multichannel cognitive radios," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4239–4253, Aug. 2010.
[16] H. Jiang, L. Lai, R. Fan, and H. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 297–307, Jan. 2009.
[17] N. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," *IEEE/ACM Trans. Netw.*, vol. 17, no. 6, pp. 1805–1818, Dec. 2009.
[18] Y. Xu *et al.*, "Optimal energy-efficient channel exploration for opportunistic spectrum usage," *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 77–80, Apr. 2012.
[19] H. Cheng and W. Zhuang, "Simple channel sensing order in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 676–688, Apr. 2011.
[20] T. Shu and H. Li, "QoS-compliant sequential channel sensing for cognitive radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2013–2025, Nov. 2014.
[21] Y. Pei, Y.-C. Liang, K. Teh, and K. H. Li, "Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, sensing order," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 1648–1659, Apr. 2011.
[22] S. Haykin, M. Fatemi, and P. Setoodeh, "Cognitive control," *Proc. IEEE*, vol. 100, no. 12, pp. 3156–3169, Dec. 2012.
[23] H. Tembine, *Distributed Strategic Learning for Wireless Engineers*. Boca Raton, FL, USA: CRC, 2012.
[24] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5667–5681, Nov. 2010.
[25] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," *Mobile Netw. Appl.*, vol. 11, no. 6, pp. 779–797, Dec. 2006.
[26] M. Felegyhazi, M. Cagalj, and J. P. Hubaux, "Efficient MAC in cognitive radio systems: A game-theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1984–1995, Apr. 2009.
[27] M. Maskery, V. Krishnamurthy, and Q. Zhao, "Decentralized dynamic spectrum access for cognitive radios: Cooperative design of a non-cooperative game," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 459–469, Feb. 2009.
[28] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games," *IEEE J. Sel. Signal Process.*, vol. 6, no. 2, pp. 180–194, Apr. 2012.
[29] H. Li, "Multi-agent Q-learning for Aloha-like spectrum access in cognitive radio systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2010, no. 1, pp. 876216-1–876216-15, May 2010.
[30] R. Fan and H. Jiang, "Channel sensing-order setting in cognitive radio network: A two-user case," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4997–5008, Nov. 2009.
[31] Y. Xu, Z. Gao, J. Wang, and Q. Wu, "Multichannel opportunistic spectrum access in fading environment using optimal stopping rule," in *Proc. ICWCA*, vol. 72, *LNICST*, 2011, pp. 275–286.
[32] Q. Wu *et al.*, "Distributed channel selection in time-varying radio environment: Interference mitigation game with uncoupled stochastic learning," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4524–4538, Nov. 2013.
[33] G. Bacci and M. Luise, "A pre-Bayesian game for CDMA power control during network association," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 76–88, Apr. 2012.
[34] L. Chen, S. Low, and J. Doyle, "Random access game and medium access control design," *IEEE/ACM Trans. Netw.*, vol. 18, no. 4, pp. 1303–1316, Aug. 2010.
[35] Q. Zhu, Z. Yuan, J. Song, Z. Han, and T. Basar, "Interference aware routing game for cognitive radio multi-hop networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 2006–2015, Nov. 2012.
[36] M. Khan, H. Tembine, and A. Vasilakos, "Game dynamics and cost of learning in heterogeneous 4G networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 198–213, Jan. 2012.
[37] T. Joshi, D. Ahuja, D. Singh, and D. P. Agrawal, "SARA: Stochastic automata rate adaptation for IEEE 802.11 networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 11, pp. 1579–1590, Nov. 2008.
[38] Y. Xing and R. Chandramouli, "Stochastic learning solution for distributed discrete power control game in wireless data networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 932–944, Aug. 2008.
[39] Y. Xing, C. Mathur, M. Haleem, and R. Chandramouli, "Dynamic spectrum access with QoS and interference temperature constraints," *IEEE Trans. Mobile Comput.*, vol. 6, no. 4, pp. 423–433, Apr. 2007.
[40] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang, "Sensing–throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.
[41] S. Hu, Y.-D. Yao, and Y. Zhuo, "Cognitive medium access control protocols for secondary users sharing a common channel with time division multiple access primary users," *Wireless Commun. Mobile Comput.*, vol. 14, no. 2, pp. 284–296, Feb. 2014.
[42] R. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, MA, USA: Harvard Univ. Press, 1991.
[43] D. Monderer and L. S. Shapley, "Potential games," *Games Econ. Behav.*, vol. 14, no. 1, pp. 124–143, May 1996.
[44] Y. Xu, Q. Wu, J. Wang, L. Shen, and A. Anpalagan, "Opportunistic spectrum access using partially overlapping channels: Graphical game and uncoupled learning," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3906–2918, Sep. 2013.
[45] Y. Xu, Q. Wu, L. Shen, J. Wang, and A. Anpalagan, "Opportunistic spectrum access with spatial reuse: Graphical game and uncoupled learning solutions," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4814–4826, Oct. 2013.
[46] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1380–1391, Apr. 2012.
[47] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. Berlin, Germany: Springer-Verlag, 2003.
[48] P. Sastry, V. Phansalkar, and M. Thathachar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 24, no. 5, pp. 769–777, May 1994.
[49] T. Börgers and R. Sarin, "Learning through reinforcement and replicator dynamics," *J. Econ. Theory*, vol. 77, no. 1, pp. 1–14, Nov. 1997.

Yuhua Xu received the B.S. degree in communications engineering and the Ph.D. degree in communications and information systems from the College of Communications Engineering, PLA University of Science and Technology, Nanjing, China, in 2006 and 2014, respectively.

Since 2012, he has been with the College of Communications Engineering, PLA University of Science and Technology, where he is currently an Assistant Professor. He has published several papers in international conferences and reputed journals in his research areas. Three of his papers are ISI highly cited papers and are ranked as core papers of Research Fronts in Opportunistic Spectrum Access. His research interests focus on opportunistic spectrum access, learning theory, game theory, and distributed optimization techniques for wireless communications.

Dr. Xu received the Certificate of Appreciation as Exemplary Reviewer for the IEEE COMMUNICATIONS LETTERS in 2011 and 2012.

Qihui Wu (SM'12) received the B.S. degree in communications engineering and the M.S. and Ph.D. degrees in communications and information systems from the Institute of Communications Engineering, Nanjing, China, in 1994, 1997, and 2000, respectively.

He is currently a Professor with PLA University of Science and Technology, Nanjing. His current research interests include algorithms and optimization for cognitive wireless networks, soft-defined radio, and wireless communication systems.

Jinlong Wang (SM'13) received the B.S. degree in mobile communications and the M.S. and Ph.D. degrees in communications engineering and information systems from the Institute of Communications Engineering, Nanjing, China, in 1983, 1986, and 1992, respectively.

Since 1979, he has been with the Institute of Communications Engineering, PLA University of Science and Technology, where he is currently a Full Professor and the Head of the Institute of Communications Engineering. He has published over 100 papers in refereed mainstream journals and reputed international conferences and has been granted over 20 patents in his research areas. His current research interests include the broad area of digital communications systems with emphasis on cooperative communication, adaptive modulation, multiple-input–multiple-output systems, soft-defined radio, cognitive radio, green wireless communications, and game theory.

Dr. Wang has served as the Founding Chair and Publication Chair of the 2009 International Conference on Wireless Communications and Signal Processing (WCSP); a member of the Steering Committee of the 2010, 2011, and 2012 WCSP; a Technical Program Committee Member for several international conferences; and a Reviewer for many prominent journals. He is currently the Vice Chair of the IEEE Communications Society Nanjing Chapter.

Liang Shen received the B.S. degree in communications engineering and the M.S. degree in communications and information system from the Institute of Communications Engineering, Nanjing, China, in 1988 and 1991, respectively.

He is currently a Professor with PLA University of Science and Technology, Nanjing. His current research interests include information theory and digital signal processing and wireless networking.

Alagan Anpalagan (M'01–SM'04) received the B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Toronto, Toronto, ON, Canada, in 1995, 1997, and 2001, respectively, all in electrical engineering.

Since August 2001, he has been with Ryerson University, Toronto, where he cofounded the WINCORE Laboratory in 2002 and currently leads the Radio Resource Management and Wireless Access and Networking R&D groups. He is currently an Associate Professor and the Program Director for Graduate Studies with the Department of Electrical and Computer Engineering, Ryerson University. He has published extensively in international conferences and journals in his research areas. His research interests include, in general, wireless communication, mobile networks, and system performance analysis, and, in particular, quality-of-service-aware radio resource management, joint study of wireless physical/link-layer characteristics, cooperative communications, cognitive radios, cross-layer resource optimization, and wireless sensor networking.

Dr. Anpalagan served as a Guest Editor for Special Issues on Radio Resource Management in 3G+ Wireless Systems (2005–2006) and Fairness in Radio Resource Management for Wireless Networks and was an Associate Editor for the EURASIP *Journal of Wireless Communications and Networking*. He previously served as the IEEE Toronto Section Chair (2006–2007), the Communications Chapter Chair (2004–2005), and the Technical Program Cochair for the IEEE Canadian Conference on Electrical and Computer Engineering (2004 and 2008). He is a Registered Professional Engineer in the province of Ontario, Canada.