Distributed Channel Selection for Interference Mitigation in Dynamic Environment: A Game-Theoretic Stochastic Learning Solution

Jianchao Zheng, *Student Member, IEEE*,
Yueming Cai, *Senior Member, IEEE*, Yuhua Xu, *Member, IEEE*, and
Alagan Anpalagan, *Senior Member, IEEE*

***Abstract*—In this paper, we investigate the problem of distributed channel selection for interference mitigation in a canonical communication network. The channel is assumed time-varying, and the active user set is considered dynamically variable due to the specific service requirement. This problem is formulated as an exact potential game, and the optimality property of the solution to this problem is first analyzed. Then, we design a low-complexity fully distributed no-regret learning algorithm for channel adaptation in a dynamic environment, where each active player can independently and automatically update its action with no information exchange. The proposed algorithm is proven to converge to a set of correlated equilibria with a probability of 1. Finally, we conduct simulations to demonstrate that the proposed algorithm achieves near-optimal performance for interference mitigation in dynamic environments.**

***Index Terms*—Distributed channel allocation, dynamic environment, interference mitigation, no-regret learning, potential game.**

I. Introduction

Efficient channel allocation plays a very important role in interference mitigation and performance improvement of communications networks. The problem of optimal channel allocation in a general network topology has been proven to be NP-hard based on its mapping to a graph-coloring problem [1]. Hence, standard optimization techniques cannot be applied directly to obtain a globally optimal solution with low computational complexity. Moreover, there is never any central control to collect the global channel state information for computing. Consequently, distributed schemes are more attractive and valuable because they require less information exchange and computational complexity and do not require a central controller [2], [3].

We can note that most existing research on distributed algorithms [4]–[7] is based on the assumptions that all the users have perfect knowledge about the environment and the actions taken by other users, and that the environment is static during the convergence of the algorithms. However, these assumptions are not realistic in practice because 1) obtaining the environment knowledge consumes a lot of network resources (e.g., time, power, and bandwidth) and may not be feasible in some emerging communication networks (e.g., *ad hoc* wireless networks and cognitive radios), and 2) the realistic channel environment is always time-varying. It can be noted that our latest work [8]–[10] address these aspects and achieve some interesting results, but there still remain some unsolved problems.

Manuscript received August 27, 2013; revised December 30, 2013 and March 3, 2014; accepted March 8, 2014. Date of publication March 12, 2014; date of current version November 6, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 61301163 and Grant 61301162 and in part by the Jiangsu Provincial Natural Science Foundation of China under Grant BK 20130067. The review of this paper was coordinated by Prof. W. Choi.

J. Zheng, Y. Cai, and Y. Xu are with the College of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China (e-mail: longxingren.zjc.s@163.com; caiym@vip.sina.com; yuhuaenator@gmail.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.2014.2311496

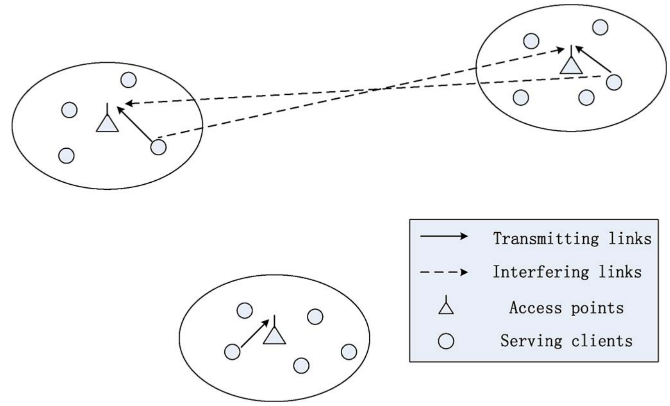

Fig. 1. Canonical network model.

In this paper, we extend our earlier work to a more general and practical system model, in which the nodes participating in the competition are variable. That is, nodes may not compete for the channels all the time due to their specific service requirements; thus, some of them may just begin to compete for the channel from a random time instance and quit the participation at a nondeterministic time as well. This case is difficult and intractable intuitively, which is the focus of this paper. Specifically, we incorporate the no-regret learning automata into the game model to solve the interference mitigation problem in a dynamic environment. The main contributions of this paper are as follows.

- We investigate the channel selection in dynamic environment when the channel and the active user set is dynamically varying, which is formulated as a stochastic dynamic game. It should be noted that the change of the active user set will lead to the change of players in the game model, which causes a large difference and intractability to the existing game framework.
- The stochastic dynamic game is proven to be an exact potential game, and the optimality property of the solution is analyzed.
- We design a low-complexity fully distributed no-regret learning algorithm to find the optimal solution in dynamic environment. The typical no-regret procedure [11] is coupled and requires a large amount of information exchange and static environment. In contrast, our proposed stochastic learning algorithm applies to dynamic environment, and each active player can independently and automatically update its action with no information exchange.

II. System Model and Problem Formulation

A. System Model

This paper studies a canonical communication network, which consists of several autonomous nodes, as shown in Fig. 1. In this network, each node is not a single communication entity but a collection of multiple entities with intranode communication capability. The entities in each collection are closely located, and there is a leading entity responsible for managing the whole collection. The leading entity chooses the operational channel, and the followers share the channel by employing some multiple-access control schemes. In [6] and [12]–[14], some instances of the canonical network, e.g., a wireless local area network access point along with its serving clients [6] and a cluster head together with its users [14], are given.

TABLE I
SUMMATION OF USED NOTATIONS

Notations	Description
$\mathcal{N}$	set of nodes
$\mathcal{M}$	set of available channels
d_{mn}	distance between nodes m and n
α	path loss exponent
β_{mn}^k	instantaneous random component of the path loss between nodes m and n in channel k
$\bar{\beta}_{mn}^k$	expected value of β_{mn}^k
H_{mn}^k	instantaneous interference gain between nodes m and n in channel k
c_n	state of node n (0 for silent, and 1 for active)
θ_n	active probability of node n
p_n	transmitting power of node n
I_n	interference experienced by node n
U	expected weighted aggregate interference
$\delta(\cdot)$	indicator function of the event in $(\cdot)$
$\mathcal{C}(t)$	active user set at time t
$\omega[t]$	a realization in the probability space at time t
$\mathcal{A}_n \equiv \mathcal{M}$	set of available actions (channels) of node (player) n
$a_n \in \mathcal{A}_n$	an action of player n
a_{-n}	an action profile of all the players except player n
u_n	expected utility function of player n
$\hat{u}_n$	state-based utility function of player n
$\Phi(a_n, a_{-n})$	potential function of the game
$\mathcal{A}$	joint strategy space
$\pi(a_n, a_{-n})$	probability distribution over joint strategy $(a_n, a_{-n}) \in \mathcal{A}$
Q_n^t	instantaneous regret matrix
R_n^t	average regret matrix
ε_t	step size of update at time t
μ	a normalization factor
z^t	empirical joint distribution of play by all the nodes
$\mathcal{C}_e$	set of correlated equilibria

In our system model, the set of nodes[1] and the set of all available channels is denoted by $\mathcal{N} = \{1, 2, \ldots, N\}, \mathcal{M} = \{1, 2, \ldots, M\}$, respectively. Assume that the interference exclusively comes from the nodes with the same channel and the leakage between different frequency bands is negligible. We also assume that all nodes are in a mutual interference area. If node m and n choose the same channel k, mutual interference emerges, and the instantaneous interference gain from nodes m to n is expressed as $H_{mn}^k = (d_{mn})^{-\alpha}\beta_{mn}^k$, where d_{mn} is the distance between nodes m and n, α is the path-loss exponent, and β_{mn}^k is the random fading coefficient. Moreover, Table I summarizes the used notations in this paper.

To study the time-varying channel environment, the channels are assumed to undergo Rayleigh fading, which is a general and realistic mobile channel model. The instantaneous random components β_{mn}^k can vary from time to time, from channel to channel, and from user to user (see [10] for detailed illustration). Additionally, in consideration of the specific service requirements for different nodes, we assume that nodes would be active/inactive with probability at each time slot. For a specific node, the active probability is stationary from the statistics perspective. We use θ_n to denote the active probability for node n. In general, the active probabilities for different nodes are different due to their different service requirements, i.e., $\theta_n \neq \theta_m$ when $n \neq m$.

B. Problem Formulation

The network utility considered in this paper is similar as in [6], [10], and [15], i.e., the expected weighted aggregate interference $U = \sum_{n\in\mathcal{N}} p_n \mathbf{E}[I_n]$, where I_n is the experienced interference by node n, $\mathbf{E}[\cdot]$ is the operation of taking expectations over the dynamic environment, and the weight of the interference experienced by node n is given by its transmission power p_n. It was shown in [15] that using such a network utility can balance the transmitting power and the experienced interference and it can lead to near-optimal network sum rate in the low-SINR regime [6]. Our goal is to find the optimal channel allocation to minimize the weighted aggregate interference when the active user set and the channel environment vary dynamically, i.e.,

$$(P1): a_{\mathrm{opt}} \in \arg\min_{a\in\mathcal{A}} U \tag{1}$$

where $\mathcal{A}$ is the joint channel allocation strategy space.

Remark 1: $P1$ is a combinatorial optimization problem and is particularly intractable under dynamic environment; hence, standard optimization techniques cannot be applied directly. Moreover, even if computational issues were to be resolved, it still requires a central controller updated with instantaneous channel gains, which would create enormous signaling overhead in practice. Therefore, designing a low-complexity fully distributed scheme to find the optimal solution is a valuable work.

III. Interference Mitigation Game

Here, the problem of distributed channel selection for interference mitigation in dynamic environment is formulated as a noncooperative stochastic dynamic game.

A. Game Model

Notably, the experienced interference is a random variable in a slot and can vary from slot to slot due to the dynamic variation of the set of players and the dynamic channel environment. Therefore, the payoffs received by players are also random in each play. We define a probability space as $(\Omega, \mathcal{H}, \mathbb{P})$, where Ω is a probability space, $\mathcal{H}$ is a minimal σ-algebra on subsets of Ω, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{H})$. Let ω denote an event in the probability space Ω. $\mathbf{X}(\omega) = [\mathbf{C}(\omega), \mathbf{H}(\omega)] : \Omega \rightarrow 2^N \times \mathbb{R}^{M\times N\times N}$ is a random vector, where $\mathbf{C} = [c_n]_{\forall n\in\mathcal{N}}$, and $\mathbf{H} = [H_{mn}^k]_{\forall m,n\in\mathcal{N},k\in\mathcal{M}}$. In our model, $c_n \in \{0, 1\}$ denotes the state of node n (0 for silent, and 1 for active), and H_{mn}^k is the channel gain between node m and node n over channel k.

For a realization $\omega[t] \in \Omega$ at time t, the state-based utility function is defined as

$$\hat{u}_n(a_n, a_{-n}, \omega[t]) = -p_n I_n(a_n, a_{-n}, \omega[t]) \tag{2}$$

where a_{-n} is a channel selection profile of all the players, excluding player n, and $I_n(a_n, a_{-n}, \omega[t])$ is the experienced interference by player n at time t. Note that $\omega[t]$ is random at different time slots. We formulate the following stochastic dynamic game as $G = [\mathcal{N}, \{\mathcal{A}_n\}_{n\in\mathcal{N}}, \{u_n\}_{n\in\mathcal{N}}]$, where $\mathcal{N}$ is the set of players, $\mathcal{A}_n$ is the set of available actions (channels) for each player n, and u_n is the expected utility function of player n, specified by $u_n(a_n, a_{-n}) = \mathbf{E}_{\mathbf{X}}[\hat{u}_n(a_n, a_{-n}, \mathbf{X})] = \lim_{T\rightarrow\infty}(1/T)\sum_{t=1}^{T}\{\hat{u}_n(a_n, a_{-n}, \omega[t])\} = -p_n\mathbf{E}[I_n]$.[2] Then, the proposed stochastic dynamic game can be expressed as

$$(G): \max_{a_n\in\mathcal{A}_n} u_n(a_n, a_{-n}), \ \forall n \in \mathcal{N}. \tag{3}$$

B. Analysis of Nash Equilibrium

Definition 1 (Nash Equilibrium): A channel selection profile $a^* = (a_1^*, a_2^*, \ldots, a_N^*)$ is a pure-strategy Nash equilibrium (NE) if and only

[1]We will use node, user, and player interchangeably in this paper.

[2]This is based on the assumption that the stochastic process is ergodic; thus, the time average of the utility function is equal to the average over the whole probability space.

if no player can improve its utility function by deviating unilaterally, i.e.,

$$u_n\left(a_n^*, a_{-n}^*\right) \geq u_n\left(a_n, a_{-n}^*\right) \qquad \forall n \in \mathcal{N}; \quad a_n \in \mathcal{A}_n. \tag{4}$$

Theorem 1: G is an exact potential game that has at least one pure-strategy NE point, and the optimal channel allocation that globally minimizes the expected weighted aggregate interference is a pure-strategy NE point of G.

Proof: First, we construct a potential function as

$$\begin{aligned}\Phi(a_n, a_{-n}) &= -\frac{1}{2}U = -\frac{1}{2}\sum_{n\in\mathcal{N}} p_n \mathbf{E}[I_n] \\ &= -\lim_{T\to\infty}\frac{1}{2T}\sum_{t=1}^{T}\sum_{n\in\mathcal{C}(t)}\sum_{m\in\{\mathcal{C}(t)\setminus\{n\}\}} p_n p_m H_{mn}^{t,a_n}\delta(a_m = a_n)\end{aligned} \tag{5}$$

where $\delta(\cdot)$ is an indicator function of the event in $(\cdot)$, and t is added as a superscript of H_{mn}^{t,a_n} to specify the time slot. $\mathcal{C}(t)$ is the active user set at time t, $\mathcal{C}(t) = \{n \in \mathcal{N} : c_n^t = 1\}$. Let $\mathcal{F}_n(a_n)$ denote the node set excluding n, which chooses a_n at time t, i.e., $\mathcal{F}_n^t(a_n) = \{m \in \{\mathcal{C}(t) \setminus \{n\}\} : a_m = a_n\}$. Therefore

$$\delta(a_n = a_m) = \begin{cases} 1, & \forall m \in \mathcal{F}_n^t(a_n) \\ 0, & \forall m \notin \mathcal{F}_n^t(a_n). \end{cases} \tag{6}$$

Then, we have

$$\begin{aligned}\Phi(a_n, a_{-n}) &= -\lim_{T\to\infty}\frac{1}{2T}\sum_{t=1}^{T}\sum_{n\in\mathcal{C}(t)}\sum_{m\in\mathcal{F}_n^t(a_n)} p_n p_m H_{mn}^{t,a_n} \\ &= -\lim_{T\to\infty}\frac{1}{2T}\sum_{t=1}^{T}\Bigg(\delta\left(n\in\mathcal{C}(t)\right)\sum_{m\in\mathcal{F}_n^t(a_n)} p_n p_m H_{mn}^{t,a_n} \\ &\qquad + \sum_{i\in\mathcal{C}(t), i\neq n}\sum_{m\in\mathcal{F}_i^t(a_i)} p_i p_m H_{mi}^{t,a_i}\Bigg) \\ &= -\lim_{T\to\infty}\frac{1}{2T}\sum_{t=1}^{T}\Bigg(\delta\left(n\in\mathcal{C}(t)\right)\sum_{m\in\mathcal{F}_n^t(a_n)} p_n p_m H_{mn}^{t,a_n} \\ &\qquad + \sum_{i\in\mathcal{C}(t), i\neq n}\sum_{m\in\mathcal{F}_i^t(a_i), m\neq n} p_i p_m H_{mi}^{t,a_i} \\ &\qquad + \delta\left(n\in\mathcal{C}(t)\right) \\ &\qquad \times \sum_{i\in\mathcal{C}(t), i\neq n} p_i p_n H_{ni}^{t,a_n}\delta(a_n = a_i)\Bigg) \\ &= -\lim_{T\to\infty}\frac{1}{2T}\sum_{t=1}^{T}\Bigg(\delta\left(n\in\mathcal{C}(t)\right)\sum_{m\in\mathcal{F}_n^t(a_n)} p_n p_m H_{mn}^{t,a_n} \\ &\qquad + \Psi_{-n} + \delta\left(n\in\mathcal{C}(t)\right) \\ &\qquad \times \sum_{m\in\mathcal{F}_n^t(a_n)} p_m p_n H_{nm}^{t,a_n}\Bigg)\end{aligned} \tag{7}$$

where $\Psi_{-n} = \sum_{i\in\mathcal{C}(t), i\neq n}\sum_{m\in\mathcal{F}_i^t(a_i), m\neq n} p_i p_m H_{mi}^{t,a_i}$ is independent of player n's strategy. Note that interference symmetry can be obtained in the canonical networks, i.e., $p_n p_m H_{mn}^{t,a_n} = p_m p_n H_{nm}^{t,a_n}$. Therefore

$$\begin{aligned}\Phi\left(a_n', a_{-n}\right) - \Phi(a_n, a_{-n}) &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Bigg(\delta\left(n\in\mathcal{C}(t)\right)\sum_{m\in\mathcal{F}_n^t(a_n')} p_n p_m H_{mn}^{t,a_n'} \\ &\qquad - \delta\left(n\in\mathcal{C}(t)\right)\sum_{m\in\mathcal{F}_n^t(a_n)} p_n p_m H_{mn}^{t,a_n}\Bigg) \\ &= u_n\left(a_n', a_{-n}'\right) - u_n(a_n, a_{-n}).\end{aligned} \tag{8}$$

According to the definition given in [16], it is known that G is an exact potential game with Φ serving as the potential function. Any global or local maxima of the potential function constitutes a pure-strategy NE point of the game G [17]. Therefore, Theorem 1 is proven. ■

Proposition 1: In underloaded or equally loaded scenarios (i.e., $N \leq M$), all pure-strategy NE points lead to interference-free channel selection profiles, which are globally optimal.

Proof: If there is a NE $a' = (a_1', a_2', \ldots, a_N')$ that is not interference free, there must exist at least two players who choose the same channel. Without loss of generality, we assume player n and m choose the same channel k, i.e., $a_n' = a_m' = k$. Obviously, it will lead to interference when player n and m are active simultaneously. Therefore, the expected interference generated by player m to player n is $\mathbf{E}[I_{m,n}] = \theta_m\theta_n p_m \mathbf{E}(H_{mn}^{t,a_n'})$. Then, we have

$$\begin{aligned}u_n\left(a_n', a_{-n}'\right) &= -p_n\mathbf{E}[I_n] \\ &\leq -p_n\mathbf{E}[I_{m,n}] \\ &= -\theta_m\theta_n p_n p_m \mathbf{E}\left(H_{mn}^{t,a_n'}\right) < 0\end{aligned} \tag{9}$$

where the first inequality holds because node n may receive interference from the other nodes, in addition to node m. Moreover, because $N \leq M$ and at least two players choose the same channel, there must exist at least one free channel. Therefore, to increase its own utility (i.e., decrease the perceived interference), player n or m will definitely deviate the current channel strategy a_n' or a_m' to choose the unoccupied free channel. Thus, a' is not a NE according to its definition, which contradicts with the former assumption. Hence, we can conclude that all NE points lead to interference-free channel selection profiles. ■

Proposition 2: If the values of expected random fading coefficients of all channels are the same, i.e., $\bar{\beta}_{mn}^k = \bar{\beta}_{mn}^0$, $\forall k \in \mathcal{M}$, then the expected aggregate interference of NE solutions in an overloaded scenario (i.e., $N > M$) that is upper bounded by $U_{\mathrm{NE}} = \sum_{n\in\mathcal{N}} p_n\mathbf{E}[I_n(a_n^*, a_{-n}^*)] \leq U_0/M$, where M is the number of channels, and $U_0 = \sum_{n\in\mathcal{N}}\sum_{m\in\{\mathcal{N}\setminus\{n\}\}}\theta_n\theta_m p_n p_m (d_{mn})^{-\alpha}\bar{\beta}_{mn}^0$ is the expected aggregate interference when all players choose the same channel.

The proof is similar to [10], which is omitted here.

IV. No-Regret Learning in Dynamic Environment

A. Algorithm Description

Here, we present a fully distributed online-adaptive no-regret learning algorithm for channel selection in dynamic environment, which is illustrated as follows.

1) Initialization: At the initial time $t = 1$, each active node $n \in \mathcal{C}(1)$ initializes its channel strategy arbitrarily.
2) Iterative update process (for $t = 2, 3, \ldots$):

- **Utility Update:** At time t, each active node $n \in \mathcal{C}(t)$ calculates the utility of the current strategy $a_n \in \mathcal{A}_n$ and the utility of choosing a different strategy $a'_n \in \mathcal{A}_n$. Then, the $|\mathcal{A}_n| \times |\mathcal{A}_n|$ instantaneous regret matrix Q_n^t is calculated by $Q_n^t(a_n, a'_n) = \delta(a_n^{(t)} = a_n) \cdot [\hat{u}_n(a'_n, a_{-n}^{(t)}, \omega[t]) - \hat{u}_n(a_n^{(t)}, a_{-n}^{(t)}, \omega[t])]$.
- **Average Regret Update**

$$D_n^t(a_n, a'_n) = D_n^{t-1}(a_n, a'_n) + \varepsilon_t \left(Q_n^t(a_n, a'_n) - D_n^{t-1}(a_n, a'_n)\right) \tag{10}$$

$$R_n^t(a_n, a'_n) = \left[D_n^t(a_n, a'_n)\right]^+ = \max\left\{D_n^t(a_n, a'_n), 0\right\} \tag{11}$$

where R_n^t represents the average regret matrix at time t, and ε_t is the step size of update.
- **Strategy Decision:** Assume a_n is the channel chosen by node n at time t, i.e., $a_n^{(t)} = a_n$. Then, at time $t+1$, node n updates its decision strategy according to the probability distribution, i.e.,

$$\begin{cases} \Pr_n^{t+1}(a'_n) = \frac{1}{\mu} R_n^t(a_n, a'_n) & \forall\, a'_n \neq a_n \\ \Pr_n^{t+1}(a_n) = 1 - \sum\limits_{a'_n \neq a_n} \Pr_n^{t+1}(a'_n) \end{cases} \tag{12}$$

where $\mu > (|\mathcal{A}_n| - 1)|u_n(a'_n, a_{-n}) - u_n(a_n, a_{-n})|\ \forall\, n \in \mathcal{N};\ \forall\, a_n, a'_n \in \mathcal{A}_n;\ \forall\, a_{-n} \in \mathcal{A}_{-n}$ is a normalization factor.

Remark 2: In the implementation of the algorithm, the values of $\hat{u}_n(a'_n, a_{-n}^{(t)}, \omega[t])$ and $\hat{u}_n(a_n^{(t)}, a_{-n}^{(t)}, \omega[t])$ are required, which depend on the channel selection strategies of other players (i.e., $a_{-n}^{(t)}$). However, it does not mean that we have to compute the utility values by obtaining the channel selection strategies of other players. Because the utility function is designed as the weighted interference as (2), $\hat{u}_n(a'_n, a_{-n}^{(t)}, \omega[t])$ and $\hat{u}_n(a_n^{(t)}, a_{-n}^{(t)}, \omega[t])$ can be achieved by measuring the interference that it experiences in each frequency band, i.e., $I_n(a'_n, a_{-n}^{(t)}, \omega[t])$ and $I_n(a_n^{(t)}, a_{-n}^{(t)}, \omega[t])$. Therefore, there is no information exchange required. The implementation of this algorithm is fully distributed among each node that can independently determine its own channel strategy.

An important contribution to the no-regret procedure is by Hart and Mas-Colell [11], but their algorithm requires a large amount of information exchange. Moreover, it requires the channel environment to be static and the player set to be fixed. The convergence of this algorithm and achievable solution in the dynamic case is an open problem.

B. Convergence Analysis

Theorem 2: Suppose that all the players perform the no-regret learning according to our proposed algorithm; then $\forall\, n \in \mathcal{N};\ \forall\, a_n, a'_n \in \mathcal{A}_n$, each player's regret $R_n^t(a_n, a'_n)$ converges to zero almost surely.

Proof: According to the algorithm, the transition probabilities are determined by the stochastic matrix defined by (12). Fix a player n and consider the Markov chain on $\mathcal{A}_n$ with transition matrix $M_n^t(a_n, a'_n) = (1/\mu) R_n^t(a_n, a'_n)$. By standard results on finite Markov chains, M_n^t admits (at least) one stationary probability measure. Let η^t be such a measure. Then (after dropping the superscript t)

$$\eta(a_n) = \sum_{a'_n \in \mathcal{A}_n} \eta(a'_n) M_n(a'_n, a_n)$$

$$= \sum_{a'_n \neq a_n} \eta(a'_n) \frac{R_n(a'_n, a_n)}{\mu} + \eta(a_n)\left(1 - \sum_{a'_n \neq a_n} \frac{R_n(a_n, a'_n)}{\mu}\right). \tag{13}$$

By collecting terms and by multiplying by μ, we have

$$\sum_{a'_n \neq a_n} \eta(a'_n) R_n(a'_n, a_n) = \eta(a_n) \sum_{a'_n \neq a_n} R_n(a_n, a'_n). \tag{14}$$

That is

$$\sum_{a'_n \neq a_n} \eta(a'_n) [D_n(a'_n, a_n)]^+ = \eta(a_n) \sum_{a'_n \neq a_n} [D_n(a_n, a'_n)]^+. \tag{15}$$

Let $\Gamma_\Omega(D_n)$ denote the projection of D_n on Ω, where Ω is the closed negative orthant of $\mathbb{R}^{|\mathcal{A}_n| \times |\mathcal{A}_n|}$. In view of [19, Prop. 3.8], it is enough to prove that inequality $\langle D_n - \Gamma_\Omega(D_n), D'_n - \Gamma_\Omega(D_n)\rangle \leq 0$. Then, by [19, Th. 5.2] and Blackwell's approachability theorem [11], we can directly get the conclusion of Theorem 4. ■

Remark 3: The proof of Blackwell's approachability theorem also gives bound on the speed of convergence [11]. Here, the speed of convergence for the expectations of the regret $E[R_n^t(a_n, a'_n)]$ is $O(1/\sqrt{(t)})$.

Definition 2 (Correlated Equilibrium): For the proposed game G, define π as the probability distribution over the joint strategy space $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_N$. The set of correlated equilibria $\mathcal{C}_e$ is the convex polytope

$$\mathcal{C}_e = \left\{\pi : \sum_{a_{-n} \in \mathcal{A}_{-n}} \pi(a_n, a_{-n}) \left[u_n(a'_n, a_{-n}) - u_n(a_n, a_{-n})\right] \leq 0,\ \forall\, n \in \mathcal{N}, a_n, a'_n \in \mathcal{A}_n\right\} \tag{16}$$

which means that, when the recommendation to player n is to choose action a_n, choosing any other action instead cannot obtain a higher expected utility.

Theorem 3: If every player follows the proposed algorithm, the empirical distributions of play z^t converge as $n \to \infty$ to the set of correlated equilibria of our game almost surely.

Proof: According to step 2 of the proposed algorithm, we have

$$D_n^t(a_n, a'_n) = \sum_{\tau \leq t: a_n^{(\tau)} = a_n} \varepsilon_{\tau-1} \left(\prod_{\lambda=\tau}^{t-1} (1 - \varepsilon_\lambda)\right) \times \left[\hat{u}_n\left(a'_n, a_{-n}^{(\tau)}, \omega[\tau]\right) - \hat{u}_n\left(a_n^{(\tau)}, a_{-n}^{(\tau)}, \omega[\tau]\right)\right]. \tag{17}$$

We set the step size to be $\varepsilon_t = 1/(t+1)$ to consider the time average (arithmetic average) in the sense of expectation; then

$$D_n^t(a_n, a'_n) = \frac{1}{t} \sum_{\tau \leq t: a_n^{(\tau)} = a_n} \left[\hat{u}_n\left(a'_n, a_{-n}^{(\tau)}, \omega[\tau]\right) - \hat{u}_n\left(a_n^{(\tau)}, a_{-n}^{(\tau)}, \omega[\tau]\right)\right]. \tag{18}$$

Let $e_j = [0, 0, \ldots, 1, 0, \ldots, 0]$ denote the $|\mathcal{A}|$ dimensional unit vector with a "1" in the jth position; thus, the empirical distribution of the

N-tuple strategy up to time t can be defined by $z^t = (1/t)\sum_{\tau\le t} e_{A^\tau}$ ($A^\tau \in \mathcal{A}$ is the joint strategy at time τ). Therefore

$$D_n^t(a_n, a_n') = \sum_{A\in\mathcal{A}:A_n=a_n} z^t(A)\left[\bar{u}_n(a_n', a_{-n}) - \bar{u}_n(a_n, a_{-n})\right] \tag{19}$$

where $\bar{u}_n(a_n, a_{-n}) = (\sum_{\tau\le t: a_{-n}^{(\tau)}=a_{-n}} \hat{u}_n(a_n, a_{-n}^{(\tau)}, \omega[\tau]))/(\sum_{\tau\le t}\delta\{a_{-n}^{(\tau)} = a_{-n}\})$ is the time average of the state-based utility function. When t is large enough, the time average is equal to the average over the entire probability space based on the assumption of ergodic property of the stochastic play, i.e.,

$$\lim_{t\to\infty}\bar{u}_n(a_n, a_{-n}) = \mathbf{E}_{\mathbf{X}}\left[\hat{u}_i(a_i, a_{-i}, \mathbf{X})\right] = u_n(a_n, a_{-n}). \tag{20}$$

Additionally, the convergence of z^t has been proven by constructing a sequence of piecewise constant continuous-time interpolated processes and then using a stochastic approximation method, i.e., $z^t \to \bar{z}$. Therefore

$$D_n^t(a_n, a_n') \to \sum_{A\in\mathcal{A}:A_n=a_n} \bar{z}(A)\left[u_n(a_n', a_{-n}) - u_n(a_n, a_{-n})\right]. \tag{21}$$

According to Theorem 2, $\lim_{t\to\infty} R_n^t(a_n, a_n') = [D_n^t(a_n, a_n')]^+ = 0$. Therefore, when $t\to\infty$, $\forall\,\alpha > 0$, $D_n^t(a_n, a_n') \le \alpha$ (obviously, $D_n^t(a_n, a_n')$ could be negative). Due to the definition of correlated equilibrium [see (16)], we can get $\lim_{t\to\infty} d(z^t, \mathcal{C}_e) = 0$, where $d(z^t, \mathcal{C}_e)$ denotes the distance between z^t and $\mathcal{C}_e$. That is, the empirical distributions of play z^t converge as $t\to\infty$ to the set of correlated equilibria of our game almost surely. ∎

It should be noted that, when the environment is static, we have $\forall\, t, \hat{u}_n(a_n, a_{-n}^{(t)}, \omega[t]) = u_n(a_n, a_{-n})$; thus, (20) also holds. Then, following the given lines of proof, we can get the theoretic results of Hart and Mas-Colell [11]. In fact, the given analysis is a generalization of the static case.

Remark 4: The set of correlated equilibria $\mathcal{C}_e$ is nonempty, closed, and convex in game G. In fact, every NE is a correlated equilibrium, and NE corresponds to the special case where the actions of the different players are independent, i.e., $\pi(a_i, a_{-i}) = \pi(a_1)\times\pi(a_2)\times\cdots\times\pi(a_N)$. Moreover, the convex set $\mathcal{C}_e$ is a convex polytope, and the NEs all lie on the boundary of the polytope [18].

C. *Computational Complexity Analysis*

At each iteration, each active node needs to keep a record of the utilities for different strategies by measuring the experienced interference. Moreover, it needs three additions and two multiplications to update one regret value, as well as one random number, one multiplication, and one comparison to select the next channel. Therefore, the computational complexity of the algorithm is $O(|\mathcal{A}_i|)$, which is low and suitable for practical implementation.

V. Simulation Results and Analysis

Here, we conduct simulations to evaluate the performance of the proposed no-regret channel adaptation algorithm for a distributed and dynamic environment.

We consider a canonical network where communication nodes are randomly scattered in a square area of 100 m × 100 m. To reflect different service requirements, the active probabilities of all nodes are randomly set in [0, 1]. Moreover, the transmitting power levels of all the nodes are randomly set in $[P_{\min}, P_{\max}]$, where $P_{\max} = 2$ W, and $P_{\min} = 1$ W. The path-loss exponent is set to be $\alpha = 2$, and the noise power experienced at each receiver is assumed identical and has a power level of -130 dBm. For simplicity, the transmitting distance for each intracommunication is set to 1 m. The number of available channels is 3, and the bandwidths of all channels are set to be 1 MHz. The Rayleigh fading model is considered in the simulation, where the channel gains are exponentially distributed with unit mean. Additionally, the normalization factor of the proposed algorithm is set to be $\mu = 10^{-3}$.

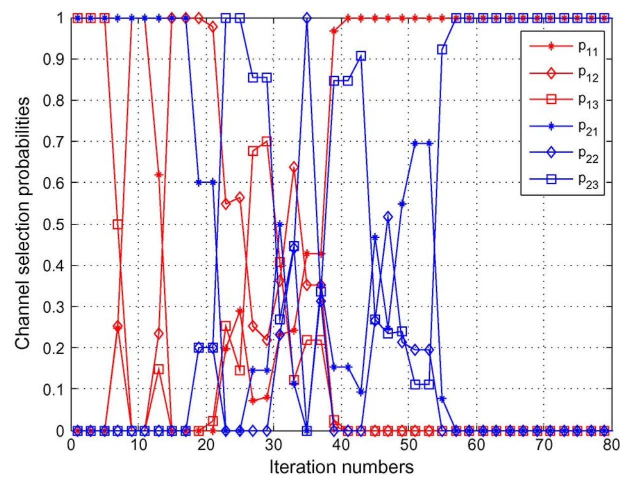

Fig. 2. Evolution of channel selection probabilities for two arbitrarily selected nodes ($N = 10$).

For convergence analysis of the proposed dynamic no-regret learning algorithm, we consider a network involving ten nodes. The convergence behavior of two arbitrarily selected nodes is shown in Fig. 2. At the beginning, each node randomly chooses a channel. As the algorithm iterates, their channel selection probabilities evolve with the time and finally converge to a pure channel strategy. Taking node 2 for example, we can see that P_{23} (the probability for choosing channel 3) converges to 1 through about 60 iterations, whereas P_{21} and P_{22} (the probability for choosing channels 1 and 2, respectively) converge to 0. Then, the channel selection probabilities remain unchanged. That is, node 2 finally chooses channel 3 by performing no-regret learning. The simulation results validate the convergence of the proposed algorithm for the interference mitigation game.

To evaluate the performance of the proposed algorithm, we additionally present the performance evaluation for a random selection scheme and the globally optimal solution for comparison. In the random selection scheme, each node randomly chooses a channel in each slot. Due to the restriction that the channel gains vary randomly and there is no information exchange, random channel selection seems to be an instinctive method. The globally optimal solution is obtained in a centralized manner when the channel characteristics and the active probability of each node are assumed known by an omnipotent genie.

Fig. 3 plots performance comparison results for the different solutions in terms of the expected weighted aggregate interference. The presented results are obtained by simulating 1000 independent trials and then by taking the expected value. Intuitively, the solution to the random channel selection scheme is the worst, which causes the most severe interference. The equilibrium solution achieved by our proposed no-regret algorithm is much better, which approaches the globally optimal solution. It is because that the learning equilibrium solution may converge to a locally/globally optimal channel selection profile as characterized by Theorem 1 and, hence, achieves near-optimal performance on average.

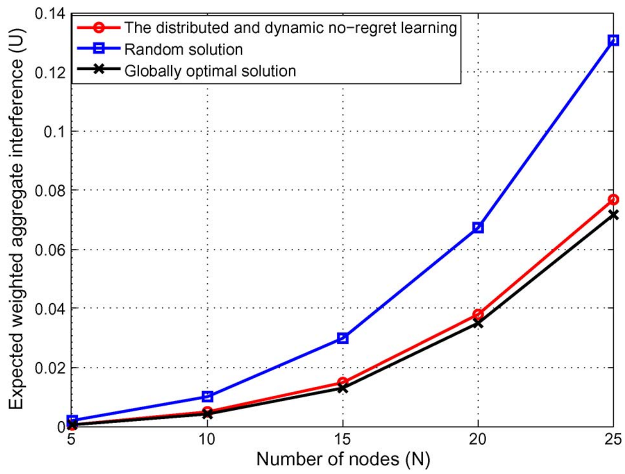

Fig. 3. Performance evaluation of expected aggregate interference for different solutions.

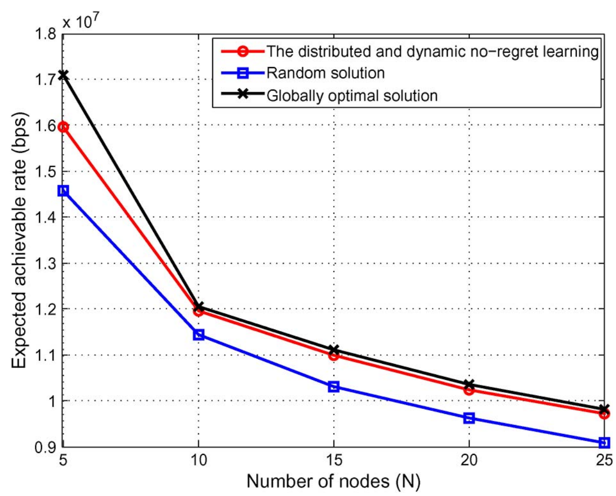

Fig. 4. Performance evaluation of expected achievable rate for different solutions.

For further illustration, we present the performance comparison in terms of the expected rate achieved by each node in Fig. 4. By the random channel selection scheme, each node achieves the worst rate gain. In contrast, our proposed dynamic no-regret learning obtains the near-optimal rate, particularly in low signal-to-interference-plus-noise ratio (SINR) cases $(N \geq 10)$, since it is proven in [6] that the minimization of weighted aggregate interference can lead to near-optimal network sum rate in a low-SINR regime. However, it may not be true in high-SINR cases. Therefore, we can see that the rate gap between the proposed algorithm and the global optimum is sharp when the number of nodes is $N < 10$, whereas the aggregate interference gap is very small, as shown in Fig. 3.

VI. Conclusion

In this paper, we have investigated the distributed channel allocation in a dynamic canonical communication network and obtained some important results. In the system model, the channel was assumed time-varying, and the active user set was considered to be dynamically variable. This problem was formulated as an exact potential game, and the optimality property of the solution was analyzed. Moreover, based on the no-regret procedure, we designed a fully distributed algorithm for dynamic channel adaptation in time-varying radio environment, where each player could independently update its action with no information exchange. The proposed algorithm exhibited low complexity and was proven to converge to a set of correlated equilibria with a probability of 1. Simulation results demonstrated the effectiveness of our proposed algorithm.

References

[1] A. Raniwala, K. Gopalan, and T. Chiueh, “Centralized channel assignment and routing algorithms for multichannel wireless mesh networks,” *ACM Mobile Comp. Commun. Rev.*, vol. 8, no. 2, pp. 50–65, Apr. 2004.

[2] Y. Xu, A. Anpalagan, Q. Wu, L. Shen, Z. Gao, and J. Wang, “Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions,” *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1689–1713, Fourth Quar., 2013.

[3] H. Zhang, L. Venturino, N. Prasad, P. Li, S. Rangarajan, and X. Wang, “Weighted sum-rate maximization in multi-cell networks via coordinated scheduling and discrete power control,” *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1214–1224, Jun. 2011.

[4] N. Nie and C. Comaniciu, “Adaptive channel allocation spectrum etiquette for cognitive radio networks,” in *Proc. IEEE DySPAN*, 2005, pp. 269–278.

[5] J. Neel and J. Reed, “Performance of distributed dynamic frequency selection schemes for interference reducing networks,” in *Proc. IEEE Milcom*, 2006, pp. 1–7.

[6] B. Babadi and V. Tarokh, “GADIA: A greedy asynchronous distributed interference avoidance algorithm,” *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6228–6252, Dec. 2010.

[7] Y. Xu, Q. Wu, J. Wang, L. Shen, and A. Anpalgan, “Opportunistic spectrum access using partially overlapping channels: Graphical game and uncoupled learning,” *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3906–2918, Sep. 2013.

[8] J. Zheng, Y. Cai, W. Yang, Y. Wei, and W. Yang, “A fully distributed algorithm for dynamic channel adaptation in canonical communication networks,” *IEEE Wireless Commun. Lett.*, vol. 2, no. 5, pp. 491–494, Oct. 2013.

[9] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y. 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.

[10] Q. Wu, Y. Xu, J. Wang, L. Shen, J. Zheng, and A. Anpalagan, “Distributed channel selection in time-varying radio environment: Interference mitigation game with uncoupled stochastic learning,” *IEEE Trans. Veh. Tech.*, vol. 62, no. 9, pp. 4524–4538, Nov. 2013.

[11] S. Hart and A. Mas-Colell, “A simple adaptive procedure leading to correlated equilibrium,” *Econometrica*, vol. 68, no. 5, pp. 1127–1150, 2000.

[12] N. Bambos, “Toward power-sensitive network architectures in wireless communications: Concepts, issues, and design aspects,” *IEEE Pers. Commun.*, vol. 5, no. 3, pp. 50–59, Jun. 1998.

[13] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y. Yao, “Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games,” *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 180–194, Apr. 2012.

[14] L. Cao and H. Zheng, “Distributed rule-regulated spectrum sharing,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 130–145, Jan. 2008.

[15] C. Lacatus and C. Popescu, “Adaptive interference avoidance for dynamic wireless systems: A game-theoretic approach,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 189–202, Jun. 2007.

[16] D. Monderer and L. S. Shapley, “Potential games,” *Games Econ. Behavior*, vol. 14, no. 1, pp. 124–143, May 1996.

[17] Y. Song, C. Zhang, and Y. Fang, “Joint channel and power allocation in wireless mesh networks: A game theoretical perspective,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1149–1159, Sep. 2008.

[18] R. Nau, S. G. Canovas, and P. Hansen, “On the geometry of Nash equilibria and correlated equilibria,” *Int. J. Game Theory*, vol. 32, no. 4, pp. 443–453, Aug. 2004.

[19] M. Benaim, J. Hofbauer, and S. Sorin, “Stochastic approximations and differential inclusions, part II: Applications,” *Math. Oper. Res.*, vol. 31, no. 3, pp. 673–695, 2006.