Distributed Learning-Based Spectrum Allocation with Noisy Observations in Cognitive Radio Networks

Mahsa Derakhshani, *Member, IEEE*, and Tho Le-Ngoc, *Fellow, IEEE*

Abstract—This paper studies the medium access design for secondary users (SUs) from a game-theoretic learning perspective. In consideration of the random return of primary users, a distributed SU access approach is presented based on an adaptive CSMA scheme, in which each SU accesses multiple idle frequency slots of a licensed frequency band with adaptive activity factors. The problem of finding optimal activity factors of SUs is formulated as a potential game, and the existence, feasibility and optimality of Nash Equilibrium (NE) are analyzed. Furthermore, to achieve NEs of the formulated game, learning-based algorithms are developed in which each SU independently adjusts its activity factors. Convergence properties of best-response dynamics and log-linear dynamics are studied. Subsequently, by learning other SUs' behavior from locally available information, the convergence with probability 1 to an arbitrarily small neighborhood of the globally optimal solution is investigated by both analysis and simulation.

I. INTRODUCTION

In hierarchical cognitive radio networks, the design of an efficient distributed medium access scheme for secondary users (SUs) is challenging due to the conflicts with licensed primary users (PUs), and hence, has received considerable attention [1]–[3]. When studying medium access design in a cognitive radio network, a key concern is how to manage spectrum sharing among SUs in a distributed manner. The simplicity, success and coexistence capability of the listen-before-talk random access scheme has the potential to be applied for the early cognitive systems [4]. For instance, Carrier Sense Multiple Access (CSMA) has been used in the literature to randomly allocate channel times among competing SUs (e.g., [5]–[7]). Furthermore, the new wireless standard IEEE 802.11af, which has been recently introduced to implement an enhanced type of Wi-Fi using TV white spaces, will also use a CSMA-based protocol for spectrum sharing [8]–[10].

Mahsa Derakhshani is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada M5S 3G4 (e-mail: mahsa.derakhshani@utoronto.ca).

Tho Le-Ngoc is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada H3A 0E9 (e-mail:tho.le-ngoc@mcgill.ca).

However, a random access scheme (e.g., CSMA) suffers from low channel utilization because of inevitable contention among users. To improve the efficiency, efforts have been exerted on optimizing random access performance (e.g., [11], [12]). Regarding IEEE 802.11 MAC operation, in [13], an opportunistic CSMA (O-CSMA) scheme is proposed for a WLAN to improve throughput by exploiting multi-user diversity gain. The proposed O-CSMA adopts a channel-adaptive back-off which prioritizes the users having high-SNR, and hence, significantly improves throughput [13]. Then, [14] proposes a multi-channel O-CSMA scheme in which high-SNR channels are selected with higher probabilities for OFDMA systems. By arranging transmission on favorable channels for multiple users, the proposed O-CSMA reaches the full multiuser diversity gain despite the distributed operation.

To offer an efficient CSMA scheme which could be compatible with requirements in a cognitive radio network, in addition to channel conditions, PU activity characteristics should be taken into consideration. Furthermore, it is important to perform fair resource allocation among SUs while avoiding collision with PUs. Thus, a more cognitive random access scheme is required to consider conflicts with PUs and fairness among SUs. Accordingly, in [15], we present an adaptive multi-channel CSMA scheme in which adaptive access probabilities (called activity factor) are adopted for SUs, taking into account PU return probabilities and spectrum sharing incentives in addition to the channel quality functions. Subsequently, the SU activity factor optimization problem for maximizing the overall SU throughput is formulated and medium access algorithms are developed for SUs.

In this paper[1], the adaptive CSMA-based access design for SUs is studied in a game-theoretic framework. There are three key reasons for using a game-theoretic approach. First, since game theory explicitly recognizes interactions among autonomous SUs, it enables providing distributed algorithms. Second, game-theoretic algorithms empower us to accelerate the convergence in comparison with the algorithm proposed in [15]. Third, game theory offers a useful tool to predict, analyze and characterize the long-run behavior of the system, specifically in comparison with the globally optimal solution.

In the proposed game-theoretic algorithm, activity factor selections are made independently and dynamically by each SU, which aims to satisfy its own demand despite imposed sharing incentives. More specifically, the activity factor optimization

---

[1]Parts of the preliminary results of this work have been presented in [16]

problem is formulated as an exact potential game. Via potential game framework, it is proved that the formulated game admits at least one pure-strategy NE, i.e., the stable operating point of the system. Subsequently, the feasibility and efficiency of NE are investigated. To address incomplete information about the game structure, learning approaches–which can respond optimally to the history information and achieve NE points–are studied in terms of information requirements and convergence properties. We establish the convergence of the best-response iterations to a pure-strategy NE which is not essentially the globally optimal solution.

Aiming to enable equilibrium selection, we introduce the *log-linear learning* process which is basically a perturbed best-response process. The log-linear learning is one of the few learning approaches which assures convergence to the most efficient NE which is globally optimal. By introducing noise into decision making process, the log-linear iterations diverge from the suboptimal NE while moving in the direction toward the globally optimal NE which is robust to noisy perturbations [17], [18].

Inspired by how the log-linear learning works, we propose a fully-distributed algorithm following best-response dynamics in which each SU adjusts its activity factors independently by learning from locally available information. Taking advantage of the noisy observations, it is shown that the best-response iterations will finally stay in a neighborhood of the globally optimal NE with probability of one. This can be explained by the fact that the globally optimal NE of the formulated game is the single stochastically stable NE. In comparison with the learning-based algorithm in [15], the proposed game-theoretic algorithm in this paper appears to have much faster convergence.

The rest of this paper is organized as follows. In Section II, we review the related work. Section III presents an overview of the system model and introduces the activity factor optimization problem. In Section IV, the medium access design is formulated as an exact potential game. Then, the existence, feasibility and efficiency of NE of the formulated game are analyzed. Section V investigates the convergence properties of learning approaches including the best-response dynamics and the log-linear dynamics. Finally, Section VI presents the concluding remarks.

II. RELATED WORK

There are several works in literature addressing distributed medium access design in hierarchical cognitive radio networks from non-game-theoretic and game-theoretic perspectives. In [5], [6], [19]–[21], the problem of channel selection is studied, assuming that both PUs and SUs have the same transmission time-slot structure. However, such an assumption of synchronous slotted transmission structure between PU and SU networks is not sensible since it needs well coordination in time between PU and SU networks. In this paper, under assumption of no time coordination between SUs and PUs, we investigate medium access design for SUs in consideration of possible collision between PU and SU due to PU return.

When studying access design under the effect of PU return, most existing works considered a single SU or a simple mechanism for random sharing among SUs without taking into account the competition or coordination among SUs [7], [22]. However, a key challenge is how to share opportunities among SUs to achieve a network-level objective. In this paper, some sharing incentives among SUs are applied to avoid channel degradation due to crowding effects. This coordination can lead to higher network performance compared to when each SU picks actions independently and competes randomly to capture the channel.

To satisfy the imposed sharing constraints in a distributed manner, a game-theoretic approach is studied and learning algorithms, including best-response dynamics and log-linear dynamics, are investigated to reach to the globally optimal solution. In the spectrum sharing literature, there are a few works studying log-linear dynamics to learn equilibrium points. For instance, [23] studies log-linear equilibriums (or logit equilibriums) for a *stochastic* game and [24] presents a log-linear learning algorithm for a *strategic* game with a *unique* pure-strategy NE. However, in this paper, log-linear learning is introduced and investigated for a *strategic potential* game with multiple pure-strategy NEs aiming to enable equilibrium selection.

Subsequently, with the aid of an adaptive CSMA scheme, the fully-distributed learning-based access algorithm is proposed in which each SU can locally and independently update its own strategy. The proposed access algorithm distinguishes this paper in coordinating spectrum access among SUs from the previously proposed channel assignment schemes for SUs with support of a central controller (e.g., [25]), exploiting a common control channel (e.g., [26]) or a simple random access (e.g., [5]).

III. SYSTEM MODEL

We study a cognitive radio network with N_s SUs looking for temporal spectrum availabilities in N_p frequency slots (or channels), licensed to PUs. It should be noted that these channels can be considered as narrowband subcarriers in a multicarrier modulation scheme (e.g., OFDM). Each SU is assumed to independently follow its *own* transmission time frame that is simply divided into equal time intervals; each interval of duration T has 2 phases: *spectrum-sensing* phase of τ followed by *spectrum-access* phase of $(T-\tau)$. Although each SU has the same values of T and τ, time synchronization is *not* necessarily required among different SUs since spectrum sharing in each channel is based on a random-access scheme, i.e., the adaptive CSMA scheme. In this paper, we assume sufficiently accurate sensing with negligible PU miss-detection, focusing on the effects of random PU returns. Perfect sensing could be a sensible assumption in certain scenarios, e.g., applications in which SUs are located inside the service area of the PU transmitter (e.g., [27]).

Let $\mathcal{N}_a := \{1,\ldots,N_a\}$ be the set of N_a channels that are detected idle in the *spectrum-sensing*, and hence, can be utilized by N_s SUs and C_k^i be the channel quality function for SU k in channel i. A simple example of the channel quality is the transmission capacity represented by $C_k^i = B^i \log(1 + P_k^i g_{k,k}^i / n_k^i)$ where B^i denotes the bandwidth, $g_{k,k}^i$

represents the channel power gain, P_k^i denotes the signal power and n_k^i represents the noise power for the SU k in the channel i. Note that we assume block flat fading in which $g_{k,k}^i$ remains unchanged during a given time-interval but independently varies from one time-interval to another.

From the SU viewpoint, the PU activity in a given channel can be modeled as a two-state continuous-time random process with the OFF (0) and ON (1) states representing the *idle* and *busy* periods of the PU. Hence, there will be a non-zero probability of PU return (i.e., α_i) during a *spectrum-access* phase to the channel i while it was detected idle by SUs in the *spectrum-sensing* phase.

Regarding spectrum sharing, it is assumed that SUs share idle channels using an adaptive CSMA approach. In the proposed CSMA, each SU k enters a competition to access an idle channel $i \in N_a$ during a *spectrum-access* phase with a certain probability (called *intended* activity factor), $\beta_k^i(0 \leq \beta_k^i \leq 1)$, to be determined based on channel qualities, PU return probabilities and sharing incentives. Adaptive activity factors enable prioritizing SUs who gain most from using a channel, and hence, improving channel utilization in comparison with a simple random access scheme. Note that β_k^i is restricted to a binary value in previous studies [5]–[7], [19]–[21], such that $\beta_k^i = 0$ expresses that the SU k does not transmit in the channel i and $\beta_k^i = 1$ represents that the SU k transmits in the channel i for the entire *spectrum-access* phase. Therefore, the proposed transmission scheme with $0 \leq \beta_k^i \leq 1$ generalizes the existing schemes.

More specifically, in the proposed CSMA scheme, each *spectrum-access* phase is divided into S equal sub-slots with length $\frac{T-\tau}{S}$, labeled $t_1, \ldots, t_S$. For every sub-slot t_j in channel i, the SU k performs the following steps:

1- Generate a Bernoulli random variable $x_k^i(t_j)$ with success probability β_k^i to be determined. If $x_k^i(t_j) = 0$, the SU k will not transmit in subslot t_j. If $x_k^i(t_j) = 1$, the SU k will proceed to the next step.
2- Generate a back-off time $\delta_k^i(t_j)$ according to a uniform distribution in the interval $(0, \delta_{\max})$.
3- After expiry of back-off time, sense the channel i, if it is idle, transmit.

In the proposed CSMA scheme, one SU with the smallest back-off time among the SUs who compete for the same sub-slot (i.e., $x_k^i(t_j) = 1$) will succeed and transmit in this sub-slot. Let $y_k^i(t_j)$ be a binary random variable representing the capturing status, i.e., $y_k^i(t_j) = 1$ if the SU k captures channel i in sub-slot t_j; otherwise, $y_k^i(t_j) = 0$.

Assuming that each SU k keeps track of its capturing status feedbacks, $y_k^i(t_j)$, the *achieved* activity factor of SU k in the channel i (i.e., the average time proportion of a *spectrum-access* phase that a SU successfully occupies the channel i, given the competition among SUs) is obtained as [15],

$$\bar{\beta}_k^i = \frac{E\left[\sum_{j=1}^{S} x_k^i(t_j) y_k^i(t_j)\right]}{S} \tag{1}$$

Then,
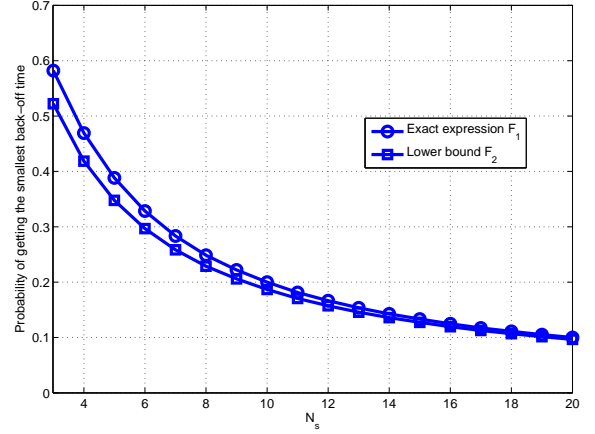
$$\bar{\beta}_k^i = P\left[y_k^i(t_j) = 1, x_k^i(t_j) = 1\right]$$

Fig. 1: Probability of getting the smallest back-off time vs. number of SUs N_s

$$= \beta_k^i P\left[y_k^i(t_j) = 1 | x_k^i(t_j) = 1\right] \tag{2}$$

where $\beta_k^i = P[x_k^i(t_j) = 1]$ represents the *intended* activity factor of SU k in the channel i. The probability of getting the smallest back-off time to capture the channel is inversely proportional to the number of SUs actively competing to capture the same channel, i.e.,

$$P\left[y_k^i(t_j) = 1 | x_k^i(t_j) = 1\right] = E\left[\frac{1}{1 + \sum_{\bar{k}=1, \bar{k} \neq k}^{N_s} x_{\bar{k}}^i(t_j)}\right] \tag{3}$$

Applying the Jensen's inequality, $P[y_k^i(t_j) = 1 | x_k^i(t_j) = 1] \geq \frac{1}{1 + \sum_{\bar{k}=1, \bar{k} \neq k}^{N_s} \beta_{\bar{k}}^i}$. Figure 1 compares $F_1 = E\left[\frac{1}{1 + \sum_{\bar{k}=1, \bar{k} \neq k}^{N_s} x_{\bar{k}}^i(t_j)}\right]$ with the lower bound $F_2 = \frac{1}{1 + \sum_{\bar{k}=1, \bar{k} \neq k}^{N_s} \beta_{\bar{k}}^i}$ for different numbers of SUs (i.e., N_s). In this example, β_k^i values are chosen randomly according to the uniform distribution in the interval $(0, 1)$. Subsequently, x_k^i is generated according to a Bernoulli random variable with success probability equal to β_k^i. It is shown that the gap between the exact expected value in F_1 and the lower bound in F_2 is very small, specifically for a larger number of SUs. Thus, this approximation does not cause any significant loss in terms of performance. Thus, in order to find a closed-form expression, $P[y_k^i(t_j) = 1 | x_k^i(t_j) = 1]$ is approximated with the lower-bound, i.e.,

$$\bar{\beta}_k^i \simeq \frac{\beta_k^i}{1 + \sum_{\bar{k}=1, \bar{k} \neq k}^{N_s} \beta_{\bar{k}}^i} \tag{4}$$

The above expression indicates that the *achieved* activity factor of each SU using the proposed CSMA scheme depends on the *intended* activity factors of other SUs.

Since the SU k in channel i transmits partially with *achieved* activity factor $\bar{\beta}_k^i$, its normalized transmission rate is $\bar{\beta}_k^i C_k^i$. Subsequently, the throughput of SU k in the idle channel i, is defined as its *successful* transmission rate $f_k^i = \bar{\beta}_k^i C_k^i(1 - \bar{\beta}_k^i \alpha_i)$ where $\bar{\beta}_k^i \alpha_i$ expresses the probability that the SU k experiences transmission loss due to PU return in idle channel

i (with probability α_i). Thus, due to the congestion nature of channel contention, the throughput of each SU is dependent on the activity factors of other SUs. To study the throughput behavior of each SU with respect to other SUs' actions, we define $s_k^i = \sum_{\bar{k}=1, \bar{k}\neq k}^{N_s} \beta_{\bar{k}}^i$ as a new parameter. Consequently, $f_k^i = \frac{\beta_k^i C_k^i}{1+s_k^i} \left(1 - \frac{\beta_k^i \alpha_i}{1+s_k^i}\right)$ where $s_k^i \geq 0$. Taking the derivative of f_k^i with respect to s_k^i, we have

$$\frac{\partial f_k^i}{\partial s_k^i} = \frac{\beta_k^i C_k^i (2\beta_k^i \alpha_i - 1 - s_k^i)}{(1+s_k^i)^3} \qquad (5)$$

Based on (5), it is clear that f_k^i reaches its maximum value at $\tilde{s}_k^i = \max(0, 2\beta_k^i \alpha_i - 1)$. Since $-1 \leq 2\beta_k^i \alpha_i - 1 \leq 1$, it can be concluded that optimal value of s_k^i happens between 0 and 1, i.e., $0 \leq \tilde{s}_k^i \leq 1$. In other words, f_k^i is a decreasing function of s_k^i for $s_k^i > 1$. This means that the achievable throughput of each individual SU decreases, while the contention among SUs is increasing (for $s_k^i \geq 1$).

Thus, to control contention among SUs and achieve higher throughput for each individual SU, we limit the sum of all activity factors of different SUs in each idle channel i to one, i.e., $\sum_{k=1}^{N_s} \beta_k^i \leq 1$, $i = 1, \ldots, N_a$. By keeping $\sum_{k=1}^{N_s} \beta_k^i \leq 1$, $i = 1, \ldots, N_a$, it can be guaranteed that $s_k^i \leq 1$, $k = 1, \ldots, N_s$, $i = 1, \ldots, N_a$. Furthermore, the sum of all activity factors of each SU k over all idle channels $i = 1, \ldots, N_a$ is set equal to one, i.e., $\sum_{i=1}^{N_a} \beta_k^i = 1$. Generally, these constraints enable us to preserve fairness among different SUs.

Aiming to determine the optimal activity factors, the optimization problem is formulated to maximize the overall throughput of all SUs under constraints of $\sum_{k=1}^{N_s} \beta_k^i = 1$, $i = 1, \ldots, N_a$. However, to reduce computational complexity, we set the utility function as $U_1 = \sum_{k=1}^{N_s} \sum_{i=1}^{N_a} \beta_k^i C_k^i (1 - \beta_k^i \alpha_i)$ while keeping $\sum_{k=1}^{N_s} \beta_k^i \leq 1$, instead of $U_2 = \sum_{k=1}^{N_s} \sum_{i=1}^{N_a} \bar{\beta}_k^i C_k^i (1 - \bar{\beta}_k^i \alpha_i)$. More specifically,

$$\max_{\boldsymbol{\beta}} \sum_{k=1}^{N_s} \sum_{i=1}^{N_a} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right) \qquad (6a)$$

$$\text{subject to } \sum_{k=1}^{N_s} \beta_k^i \leq 1, \ i = 1, \ldots, N_a \qquad (6b)$$

$$\sum_{i=1}^{N_a} \beta_k^i = 1, \ k = 1, \ldots, N_s \qquad (6c)$$

$$0 \leq \beta_k^i \leq 1, \ i = 1, \ldots, N_a, \ k = 1, \ldots, N_s \qquad (6d)$$

The formulated activity factor optimization problem is feasible if $N_s \leq N_a$. Note that this paper introduces the *successful* transmission rate to establish the utility function of each SU in each channel. However, most existing works, e.g., [5], [6], [19]–[21], ignore the effects of random PU returns on the SU transmission and the utility function of each SU in each channel is defined as its transmission rate. Furthermore, in previous studies, [5], [6], [19]–[21], the transmission rate of each SU in each channel is simply defined as the channel bandwidth assuming that number of bits delivered over a channel in one time-slot is proportional to its bandwidth. Instead, in this paper, we consider the channel quality function which reflects channel conditions as well.
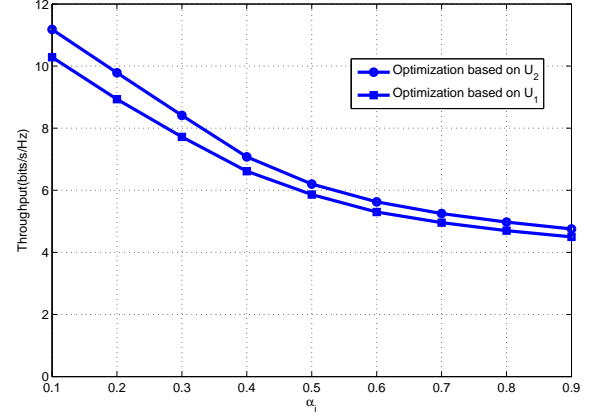

Fig. 2: Overall throughput of SUs vs. the primary returning probability, α_i, for fixed $N_s = 3$, $N_a = 3$ and SNR = 10dB
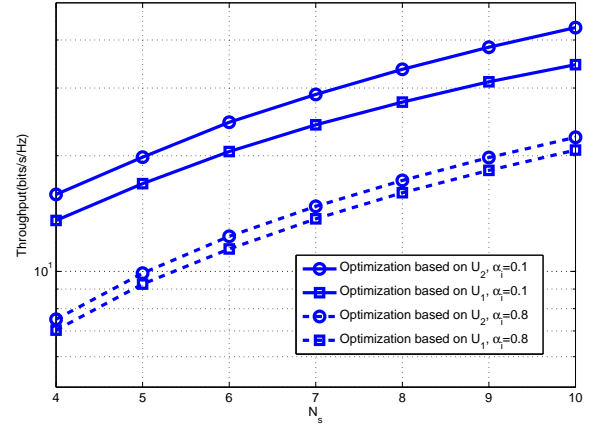

Fig. 3: Overall throughput of SUs vs. the number of SUs, N_s, for fixed $\alpha_i = 0.1$, $\alpha_i = 0.8$ and SNR = 10dB

In order to investigate the effect of choosing U_1 along with $\sum_{k=1}^{N_s} \beta_k^i \leq 1$ instead of U_2 to find optimal activity factors, we plot the overall throughput of all SUs (i.e., $\sum_{k=1}^{N_s} \sum_{i=1}^{N_a} \bar{\beta}_k^i C_k^i (1 - \bar{\beta}_k^i \alpha_i)$) versus α_i and N_s for both scenarios (i.e., using U_1 and U_2) in Figures 2 and 3. In this example, we assume independent channels with the same bandwidth $B^i = 1$ and same α_i. Please note that we used exhaustive search to find the optimal activity factors while optimizing U_2. It is shown that the throughput improvement by using U_2 instead of U_1 (along with $\sum_{k=1}^{N_s} \beta_k^i \leq 1$) to find optimal activity factors is small. This fact can be explained by the effective role of keeping $\sum_{k=1}^{N_s} \beta_k^i \leq 1$ to reduce competition among SUs. On the other hand, using U_2 makes the optimization problem non-convex and computationally complex. It is worth mentioning that we have studied the activity factor allocation problem based on U_2 in [28]. Since the optimization problem is generally non-convex, the medium access design is cast in a game theoretic framework. It is shown that the NE of the formulated game can be highly inefficient in terms of total profit for all SUs. However, by setting the utility function as $\sum_{k=1}^{N_s} \sum_{i=1}^{N_a} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right)$

while keeping $\sum_{k=1}^{N_{\mathrm{s}}} \beta_k^i \leq 1$, in this paper, it will be shown that the globally optimal solution is achievable through a fully-distributed game-theoretic algorithm.

IV. GAME-THEORETIC DESIGN

In a highly dynamic environment such as cognitive radio networks, it is practically essential to find a reasonably good solution which can be obtained in a sufficiently fast manner. To this end, in this section, the CSMA-based design is studied in a game-theoretic framework which enables distributed implementation and fast convergence to a reasonably good solution. More specifically, we consider a strategic non-cooperative game in which the players are SUs.

According to the optimization problem in (6), each SU could simply maximize its transmission rate (i.e., $\sum_{i=1}^{N_{\mathrm{a}}} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right)$). However, SUs cannot select activity factors which violate the coupling constraints in (6b). Since it is difficult for SUs to identify feasible activity factors in advance, we consider a penalty mechanism embedded in the payoff function of each SU as a cost function. The penalty mechanism is aimed to satisfy the coupling constraints among different SUs in (6b) by severely punishing the SU who violates each of them. Specifically, we focus on a linear penalty mechanism. Thus, we construct an alternative payoff function for SU k as

$$u_k = \sum_{i=1}^{N_{\mathrm{a}}} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right) - \sum_{i=1}^{N_{\mathrm{a}}} \mu_i \Theta \left(\sum_{j=1}^{N_{\mathrm{s}}} \beta_j^i - 1\right) \quad (7)$$

where $\Theta(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$ and μ_i is the penalty factor in channel i which is a positive scalar. The second term of (7) represents the penalty mechanism in the payoff function of each SU. More specifically, we define the penalty in channel i as $\mu_i \Theta(\sum_{k=1}^{N_{\mathrm{s}}} \beta_k^i - 1) = \begin{cases} \mu_i(\sum_{k=1}^{N_{\mathrm{s}}} \beta_k^i - 1), & \sum_{k=1}^{N_{\mathrm{s}}} \beta_k^i > 1 \\ 0, & \sum_{k=1}^{N_{\mathrm{s}}} \beta_k^i \leq 1 \end{cases}$. In other words, on one hand, SU k will not be penalized in channel i if β_k^i is chosen such that $\sum_{k=1}^{N_{\mathrm{s}}} \beta_k^i \leq 1$. On the other hand, SU k has to pay a penalty equal to $\mu_i(\sum_{k=1}^{N_{\mathrm{s}}} \beta_k^i - 1)$ if β_k^i is chosen such that $\sum_{k=1}^{N_{\mathrm{s}}} \beta_k^i > 1$. Note that SUs need to use the same values of penalty factors, which can be easily set up during initialization.

Let $\boldsymbol{\beta}_k = \left[\beta_k^1, \ldots, \beta_k^{N_{\mathrm{a}}}\right]$ be the pure strategy at SU k and $\boldsymbol{\beta}_{-k}$ be the pure strategy of all SUs excluding the SU k. A pure strategy defines a specific action that a player can pick from its own strategy set. Such actions are not drawn from a probability distribution, as in the case of mixed strategies. Subsequently, the admissible pure strategies of SU k is defined as

$$\mathcal{B}_k = \left\{\boldsymbol{\beta}_k : \beta_k^i \in \{0, \frac{1}{S}, \frac{2}{S}, \ldots, 1\}, \forall i \in \mathcal{N}_a, \sum_{i=1}^{N_{\mathrm{a}}} \beta_k^i = 1\right\} \quad (8)$$

Note that β_k^i takes discrete values in the proposed CSMA-based algorithm. Then, we can define the non-cooperative

game for the medium access of SUs as

$$g = [\mathcal{N}_s, \{\mathcal{B}_k\}_{k \in \mathcal{N}_s}, \{u_k\}_{k \in \mathcal{N}_s}] \quad (9)$$

where $\mathcal{N}_s = \{1, \ldots, N_s\}$ is the set of players of the game (i.e. SUs), $\mathcal{B}_k$ is the activity factor strategy set of the SU k and u_k is the corresponding payoff function of the SU k defined on the set of pure-strategy profiles $\mathcal{B} = \mathcal{B}_1 \times \ldots \times \mathcal{B}_{N_s}$.

An exact potential game is a strategic game in which the incentive of all players to change their strategies can be expressed in a global potential function. The potential games are easy to analyze since improving each player's utility also increases the value of a potential function [29]. In the following theorem, we demonstrate that the game $\mathcal{G}$ falls into the framework of exact potential games.

Theorem 1: The formulated game $\mathcal{G}$ is an exact potential game with the potential function,

$$\Phi = \sum_{j=1}^{N_s} \sum_{i=1}^{N_a} \beta_j^i C_j^i (1 - \beta_j^i \alpha_i) - \sum_{i=1}^{N_a} \mu_i \Theta \left(\sum_{j=1}^{Ns} \beta_j^i - 1\right) \quad (10)$$

Proof: It is clear that the game $\mathcal{G}$ satisfies the exact potential game definition [29],

$$u_k(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}) - u_k(\boldsymbol{\beta}_k', \boldsymbol{\beta}_{-k}) = \Phi(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}) - \Phi(\boldsymbol{\beta}_k', \boldsymbol{\beta}_{-k}),$$
$$\forall \boldsymbol{\beta}_k, \boldsymbol{\beta}_k' \in \mathcal{B}_k, \forall k \in \mathcal{N}_s \quad (11)$$

Thus, $\mathcal{G}$ is an exact potential game and Φ is the potential function of $\mathcal{G}$. ∎

Conceptually, a strategic game can reach a steady-state Nash Equilibrium (NE) point, if it exists, from which no player can improve its utility by changing its own strategy unilaterally [30]. In other words, a pure-strategy profile $\boldsymbol{\beta}^* = \{\boldsymbol{\beta}_k^*\}_{k=1}^{N_s} \in \mathcal{B}$ is a pure-strategy NE if and only if

$$u_k(\boldsymbol{\beta}_k^*, \boldsymbol{\beta}_{-k}^*) \geq u_k(\boldsymbol{\beta}_k', \boldsymbol{\beta}_{-k}^*), \forall \boldsymbol{\beta}_k' \in \mathcal{B}_k, \forall k \in \mathcal{N}_s \quad (12)$$

We are interested to investigate the existence and characteristics including the feasibility and efficiency of NE of the game $\mathcal{G}$.

A. Existence

First of all, the existence of NE of the game $\mathcal{G}$ is studied in the following theorem based on the properties of the potential games.

Theorem 2: The game $\mathcal{G}$ admits at least one pure-strategy NE.

Proof: This theorem comes directly from Corollary 4 in [31], which states every finite potential game $\mathcal{G}$ has at least one pure-strategy NE. ∎

It should be noted that the payoff function of each SU (i.e., u_k) in (7) is not a continuous function over the strategy set $\mathcal{B}$. Thus, payoff functions are not strictly concave. As a result, it cannot be concluded that the pure-strategy NE of the formulated game $\mathcal{G}$ is always unique.

Remark 1: In general, the pure-strategy NE of game $\mathcal{G}$ may not be unique.

B. Feasibility

Since the optimization problem in (6) has the coupling constraints in (6b) which are merged in the payoff functions in the formulated game $\mathcal{G}$, it is required to verify if an arbitrary pure-strategy NE is feasible, i.e., satisfying the constraints $\sum_{k=1}^{N_s} \beta_k^i \leq 1, i = 1, \ldots, N_a$. Thus, the following theorem presents conditions that assure the feasibility of pure-strategy NEs.

Theorem 3: All pure-strategy NEs of the game $\mathcal{G}$ must be feasible if

$$\mu_i > \mu_{th}, \forall i \in \mathcal{N}_a \qquad (13)$$

where

$$\mu_{th} = \max_{k \in \mathcal{N}_s, \boldsymbol{\beta}_k \in \mathcal{B}_k} \left(S \sum_{i=1}^{N_a} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right) \right) \qquad (14)$$

Proof: See Appendix A for the proof.

Theorem 3 assures that, by properly designing μ_i's, the payoff functions in (7) can guarantee the feasibility of the steady states of the system.

C. Efficiency

The other aspect that we are interested to study is how efficient the NE of game $\mathcal{G}$ is in comparison with the globally optimal solution of (6). In other words, NE efficiency is measured comparing to the globally optimal solution of (6). The following theorem specifies the relationship between the globally optimal solution and the NE of the game $\mathcal{G}$.

Theorem 4: The globally optimal solution of (6) is the Pareto-optimal pure-strategy NE of $\mathcal{G}$ if $\mu_i > \mu_{th}, \forall i \in \mathcal{N}_a$.

Proof: See Appendix B for the proof.

Theorem 4 shows that the globally optimal solution constitutes a pure-strategy NE. From Theorem 4, it follows that the NE must be the globally optimal solution when the NE is unique. However, when multiple NEs exist in the formulated game $\mathcal{G}$, a randomly chosen NE may not be the globally optimal solution. That is why we need to study different learning algorithms to find an iterative algorithm which can guarantee convergence to the most efficient NE (i.e., the globally optimal solution).

Remark 2: In the next section, learning-based iterative algorithms are proposed which enable the convergence to the globally optimal pure-strategy NE.

V. LEARNING EQUILIBRIA

Assuming that the rationality of players and the structure of the game are common knowledge, equilibrium can be observed as a result of analysis and introspection of the players. Otherwise, under assumption of bounded rationality or partial information, equilibrium may arise as a consequence of a long-run learning process. In this section, aiming to achieve an equilibrium of the game $\mathcal{G}$, learning approaches are discussed in terms of information requirements and convergence properties.

A. Best-Response Dynamics

To reach an equilibrium of the game $\mathcal{G}$, first, a simple learning algorithm for activity factor selection is presented, based on the asynchronous best-response dynamics. In particular, at each time $t \in \{0, 1, 2, \ldots\}$, exactly one SU $k \in \mathcal{N}_s$, is randomly selected to revise its activity factors based on the best-response dynamics defined as

$$\boldsymbol{\beta}_k[t] = \arg \max_{\boldsymbol{\beta}_k' \in \mathcal{B}_k} u_k(\boldsymbol{\beta}_k', \boldsymbol{\beta}_{-k}[t-1]) \qquad (15)$$

Under perfect knowledge of current strategies of other SUs (i.e., $\boldsymbol{\beta}_{-k}[t-1]$), the convergence of the proposed game-theoretic algorithm is established in the following theorem.

Theorem 5: The learning algorithm under asynchronous best-response dynamics converges with probability of one to a pure-strategy NE of the game $\mathcal{G}$ from any initial strategy point.

Proof: Based on Theorem 19 in [30], in a finite exact potential game, best-response dynamics will converge with probability of one to a pure-strategy NE in finite steps. Accordingly, in game $\mathcal{G}$, the best-response iterations will converge to a pure-strategy NE. ∎

To verify convergence of the best-response dynamics, a numerical result is also provided. In this example, we assume independent channels with the same bandwidth $B^i = 1$ and the same α_i. We set the similar SNR $= \frac{P_k^i}{n_k^i} = 10$dB for individual SUs. Furthermore, we assume $N_s = 3$, $N_a = 3$ and $\alpha_i = 0.1$. Note that we use the same example for all simulation results.

Figure 4 demonstrates the convergence process of the activity factors of three SUs in the first idle licensed channel for a certain channel realization. All SUs start by setting their initial values of their activity factors to zero. In each iteration, they sequentially play to optimize their own payoff functions, and update their activity factors based on the best-response dynamics in (15) (e.g., in the following order: SU3, SU2, SU1). Note that each SU updates its activity factors in three different channels simultaneously. However, due to the limited space, the simulation results for only channel 1 are demonstrated.

In iterations 1, SU3 then SU2 then SU1 play and obtain $\beta_3^1[1] = 0.5$, $\beta_2^1[1] = 0.33$, and $\beta_1^1[1] = 0.53$, which result in $\sum_{k=1}^{3} \beta_k^1 \geq 1$. In iteration 2, SU3 is penalized for this excessive amount, and, for its specific channel realization, Figure 4 indicates that the SU 3 is forced to reduce its activity factor to $\beta_3^1[2] = 0.14$ in order to keep $\sum_{k=1}^{3} \beta_k^1 \leq 1$. Subsequently, SU2 and SU1 do not need to change their strategies, and maintain $\beta_2^1[2] = 0.33$, and $\beta_1^1[2] = 0.53$. From iteration 3, since $\sum_{k=1}^{3} \beta_k^1 \leq 1, i = 1, \ldots, 3$, SUs actually optimize their own throughput, i.e., their own payoff functions become $u_k = \sum_{i=1}^{3} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right)$. Thus, their strategies do not change any more. Note that, for different channel realizations, it is possible that all SUs change their activity factors to satisfy $\sum_{k=1}^{N_s} \beta_k^i \leq 1$ in iteration 2. Furthermore, the process of reducing $\sum_{k=1}^{N_s} \beta_k^i$ may take more than two iterations. Additionally, Figure 5 shows the convergence of potential function. As evident from Figures 4 and 5, in less than 10 iterations, activity factor selection converges to a NE
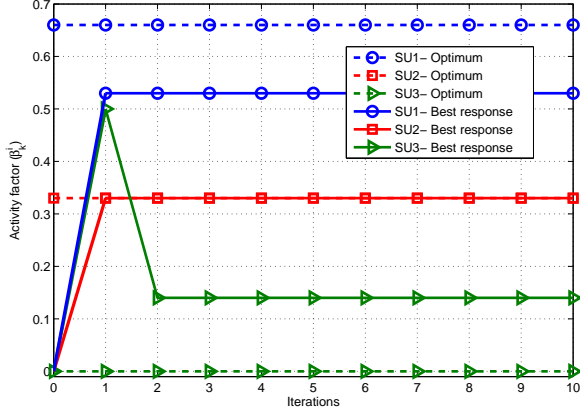
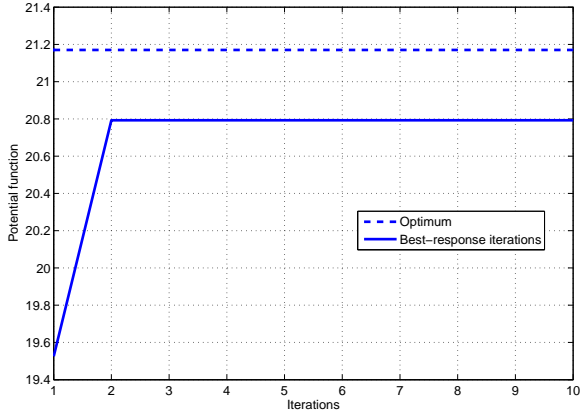Fig. 4: Convergence of the SU activity factors with best-response dynamics.

Fig. 5: Convergence of the potential function with best-response dynamics.

which is not essentially the globally optimal solution. Note that each iteration corresponds to a complete update by all the SUs.

According to Theorem 5, the best-response iterations will converge to a pure-strategy NE which is not necessarily the globally optimal solution of (6). Since NE could be highly inefficient with regards to the network-level objective, it is thus crucial to find a learning process to reach the globally optimal solution.

B. Log-Linear Dynamics

Since the basic myopic best-response dynamics suffer from multiple rest points (e.g., any pure-strategy NE), it is essential to introduce a learning process that can select an appropriate equilibrium. Aiming to enable equilibrium selection, the log-linear learning has been proposed as a perturbed best-response process which guarantees convergence to the most efficient NE for potential games [17], [18]. The basic idea behind the equilibrium selection in log-linear learning is to introduce noise into decision making process which enables categorizing equilibria based on their stability characteristics. This noise allows players to select suboptimal actions (i.e., non-best

responses) with a certain probability which is attached with the magnitude of the payoff difference of the best response and the suboptimal action.

In the log-linear learning, SUs are assumed to be myopic and boundedly rational. At each time $t > 0$, exactly one SU $k \in \mathcal{N}_s$ is randomly selected to update its action, using a probability distribution over its strategy set in response to the current strategy profile. Let k be the player chosen at time t to revise its action. Then, SU k will choose action $\boldsymbol{\beta}_k$ given the current strategy profile $\boldsymbol{\beta}_{-k}[t-1]$ with a probability based on the log-linear choice rule

$$\pi_k^{\boldsymbol{\beta}_k}[t] = \frac{e^{\frac{1}{\epsilon} u_k(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}[t-1])}}{\sum_{\boldsymbol{\beta}'_k \in \mathcal{B}_k} e^{\frac{1}{\epsilon} u_k(\boldsymbol{\beta}'_k, \boldsymbol{\beta}_{-k}[t-1])}} \quad (16)$$

where $0 < \frac{1}{\epsilon} < \infty$. The scalar $\frac{1}{\epsilon}$ can be interpreted as the level of rationality of the SUs. In other words, ϵ shows the level of noise in the SUs' decisions and determines how often SUs choose their best responses. The described rule is called log-linear since the log-likelihood ratio of selecting between two actions is linearly proportional to the difference of corresponding payoffs, given other SUs' actions [32]. As $\epsilon \to 0$, the log-linear rule approaches to the best-response rule. However, the SU k will choose any action $\beta_k \in \mathcal{B}_k$ with equal probability as $\epsilon \to \infty$. Thus, for any $0 < \epsilon < \infty$, SUs explore non-best responses with non-zero probabilities which are exponentially smaller for actions yielding smaller payoffs [17], [18].

Our goal is to characterize the long-run behavior of asynchronous log-linear learning process for game $\mathcal{G}$. To this end, the log-linear dynamic adjustment process is represented as an irreducible and aperiodic Markov chain $\{X_t^\epsilon\}_{t \in \mathbb{N}}$ on the set of strategy profiles of the game [17], [18]. Subsequently, based on the study in [17], the stationary distribution, i.e., limiting distribution in a Markov chain, is studied to explain the long-run behavior of this update rule. Accordingly, the stationary distribution of asynchronous log-linear learning process in game $\mathcal{G}$ is presented as

$$\mu^\epsilon(\boldsymbol{\beta}) = \frac{e^{\frac{1}{\epsilon} \Phi(\boldsymbol{\beta})}}{\sum_{\boldsymbol{\beta} \in \mathcal{B}} e^{\frac{1}{\epsilon} \Phi(\boldsymbol{\beta})}} \quad (17)$$

According to (17), $\mu^\epsilon(\boldsymbol{\beta})$ (i.e., the probability that $X_t^\epsilon = \boldsymbol{\beta}$ for sufficiently large times $t > 0$) can be expressed as an explicit function of the potential function. A strategy profile $\boldsymbol{\beta} = \{\boldsymbol{\beta}_k\}_{k=1}^{N_s}$ is said to be stochastically stable if $\lim_{\epsilon \to 0} \mu(\boldsymbol{\beta}) > 0$. Consequently, the stochastically stable strategy can be computed based on (17).

Corollary 1: In game $\mathcal{G}$, the only stochastically stable strategy profile of asynchronous log-linear learning process is the maximizer of potential function Φ which is the NE of the potential game $\mathcal{G}$ based on Theorem 2 in [29]. Assuming $\mu_i > \mu_{th}, \forall i \in \mathcal{N}_a$, the globally optimal solution of (6) is equal to the maximizer of the potential function. As a result, the globally optimal solution of (6) is the only stochastically stable NE.

The importance of this Corollary–which comes directly from Corollary 1 in [17] for exact potential games–is two-fold, i.e., convergence and equilibrium selection. In particular, this
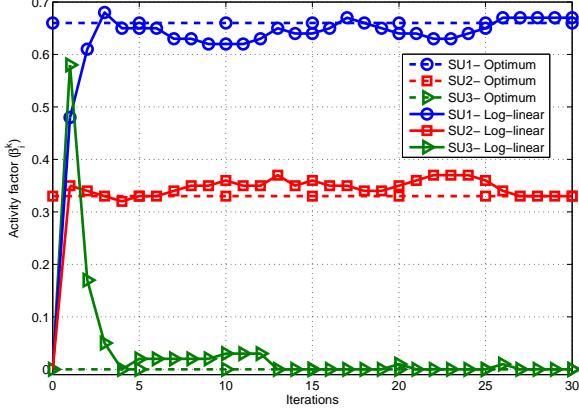
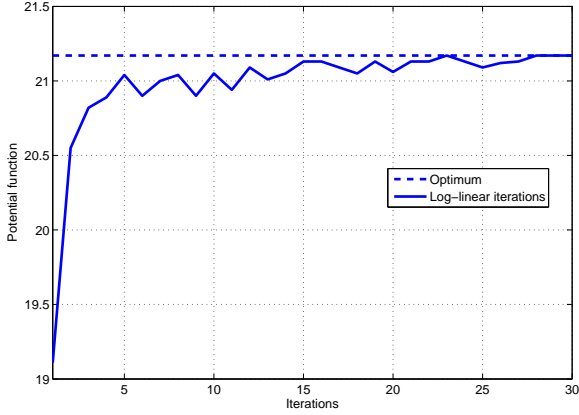Fig. 6: Convergence of the SU activity factors with log-linear dynamics.

Fig. 7: Convergence of the potential function with log-linear dynamics.

explains that the log-linear learning converges, in probability, to the NE, which is the potential function maximizer (and also equal to the globally optimal solution in the formulated game $\mathcal{G}$ assuming $\mu_i > \mu_{th}, \forall i \in \mathcal{N}_a$) with sufficiently large rationality level. In other words, by properly designing the rationality level, log-linear dynamics will ultimately converge to the globally optimal NE in game $\mathcal{G}$.

To verify the convergence properties of log-linear learning, Figures 6 and 7 show the convergence process of the activity factors of three SUs in the first idle licensed channel for a certain channel realization and the convergence of the potential function, respectively. They confirm that the log-linear iterations lead to the globally optimal solution despite the best-response dynamics. As can be observed, it takes merely around 20 iterations to quickly converge to the globally optimal solution. The other aspect which affects the practicality of a learning process is the convergence speed. From Proposition 152 in [33], assuming $\mu_i > \mu_{th}$, the convergence time of the log-linear learning to be η-close to the globally optimal solution is in the order of

$$N_s \log \log(N_s) + \log\left(\frac{1}{\eta}\right) \tag{18}$$

for any initial condition if the rationality level (i.e., $\frac{1}{\epsilon}$) is sufficiently large. According to (18), the convergence time is linearly proportional to the number of SUs using the log-linear learning in game $\mathcal{G}$.

Note that, in the proposed adaptive CSMA scheme, a player (SU) updates its strategy for its iterative learning algorithm (using best-response or log-linear dynamics) after a successful transmission and does not need any coordination with other SUs.

Considering the proposed best-response dynamics and log-linear dynamics, each SU (e.g., SU k) needs the sum of current activity factors of all other SUs in each channel (i.e., $s_k^i[t] = \sum_{\bar{k}=1, \bar{k} \neq k}^{N_s} \beta_{\bar{k}}^i[t-1], \ i = 1, \ldots, N_a$) to update its own strategy. Up to this point, we assume that SU k can obtain $s_k^i[t]$ accurately, with the aid of a central coordinator or an information exchange between SUs, i.e., SU k has perfect knowledge of $s_k^i[t], i = 1, \ldots, N_a$.

In the next section, to relief the need for frequent information exchange in the proposed learning algorithms, an alternative algorithm has been proposed based on best-response dynamics using locally available information. In the proposed algorithm, SU k estimates $s_k^i[t]$ based on the CSMA feedbacks. Since this estimation with limited samples suffers from random errors, each SU would have only noisy estimation of $s_k^i[t]$ in the proposed algorithm. Then, considering such noisy information, convergence of the proposed iterative algorithm following best-response dynamics is studied. Since each SU can locally and independently update its own strategy without any information exchange among SUs, this proposed iterative algorithm can be implemented in a distributed manner.

VI. BEST-RESPONSE DYNAMICS WITH NOISY OBSERVATIONS

In [15], we study how to use the capturing status feedbacks of CSMA scheme, $y_k^i(t_j)$, to estimate the sum of activity factors of all SUs in each channel. It is shown that $\beta^i = \sum_{k=1}^{N_s} \beta_k^i$ can be updated after each window of S' sub-slots as $\hat{\beta}^i \simeq \beta_k^i + \left(\frac{S'.\beta_k^i}{\sum_{l=f+1}^{f+S'} y_k^i(t_l)}\right) - 1$. Since estimation with a limited number of samples suffers from random errors, it is shown that $\hat{\beta}^i = \beta^i + w$ where $\mathrm{E}[w]$ and $\mathrm{var}[w]$ are of $\mathcal{O}\left((S')^{-1}\right)$ [15].

From (7), it is clear that the estimation noise of β^i will cause a bias (i.e., $b = \max(0, w)$) in u_k, and hence, best-response iterations in (15) will also involve random errors. Since the first derivative of u_k is finite, the bias and variance of the random noise in best-response iterations should be also of $\mathcal{O}\left((S')^{-1}\right)$.

As shown in log-linear learning, adding noise to the decision making process enables equilibrium selection taking advantage of the fact that the globally optimal NE is the only stochastically stable NE in potential games. Accordingly, in Theorem 3 of [34], it is shown that a bounded noise will asymptotically ensure the convergence of the best-response iterations to a neighborhood of the globally optimal solution in potential games. That is because suboptimal NE points are less stable than the globally optimal NE (i.e., global optimum) in a
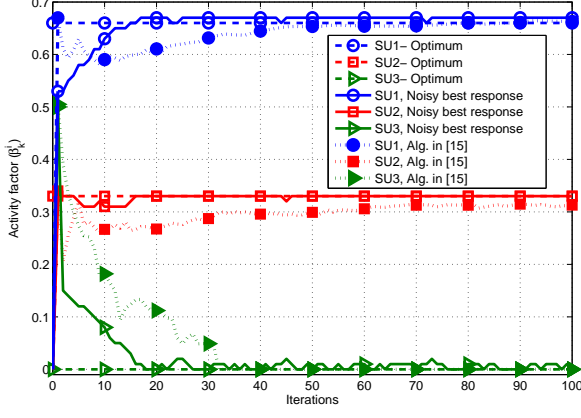
Fig. 8: Convergence of the SU activity factors with noisy best-response dynamics.

Fig. 9: Convergence of the potential function with noisy best-response dynamics.

sense that a small noise can cause the best-response iterations diverge from the suboptimal NE while moving in the direction toward the globally optimal NE.

Similarly, in the proposed algorithm, best-response iterations involve errors although they are random with bounded bias and variance. With a sufficiently large or increasing estimation window (i.e., S'), the random noise can be approximated as a bounded noise. Therefore, it is expected that the best-response iterations converge to the global optimum. Mathematically, this can be presented as the following claim.

Claim 1: $\forall \varepsilon > 0$, an estimation window size can be selected (i.e., $\exists S' > 0$) such that $\lim_{t \to \infty} \inf \Phi(\boldsymbol{\beta}[t]) \geq \Phi_{\max} - \varepsilon$ with probability of 1.

This claim declares that, by properly designing an estimation window size (i.e., S'), best-response iterations can get arbitrarily close to the globally optimal solution of (6) which is also the maximizer of the potential function (i.e., Φ), assuming $\mu_i > \mu_{th}, \forall i \in \mathcal{N}_a$.

To confirm convergence of noisy best-response iterations to the globally optimal solution, Figure 8 demonstrates the convergence process of the activity factors of three SUs in the first idle licensed channel. In addition, the convergence process based on the proposed learning-based non-game-theoretic algorithm in [15] is illustrated in Figure 8. It is clear that the game-theoretic algorithm accelerate the convergence in comparison with the algorithm proposed in [15]. Furthermore, Figure 9 shows the convergence of the potential function. They confirm that the best-response iterations will stay in a neighborhood of the global optimum.

VII. CONCLUSIONS

In this paper, using a potential game model, we developed a distributed adaptive CSMA-based access design in which activity factors of SUs over multiple idle channels are adaptively adjusted. Via potential game framework, it is established that the formulated game admits at least one pure-strategy NE. In consideration of coupled constraints among SUs, sufficient
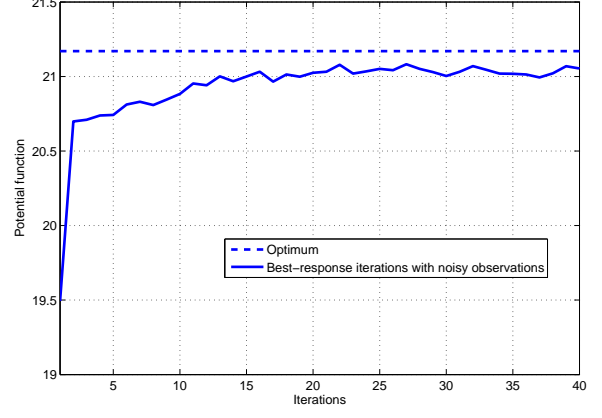
conditions are presented to assure the feasibility of pure-strategy NEs. In addition, it is proved that the globally optimal solution is the Pareto-optimal NE.

Furthermore, the convergence properties of the best-response dynamics and log-linear dynamics of the formulated game are investigated. Assuming perfect knowledge of moves previously made by all SUs, it is proved that best-response iterations converge to a pure-strategy NE which is not essentially the globally optimal solution. However, the log-linear process enables equilibrium refinement and convergence to the most socially desirable solution. Subsequently, in a game with noisy observations, it is shown that best-response iterations also converge with probability of one to a neighborhood of the global optimum.

APPENDIX

A. Proof of Theorem 3

Suppose that $\boldsymbol{\beta} = \{\boldsymbol{\beta}_k\}_{k=1}^{N_s}$, where $\boldsymbol{\beta}_k \in \mathcal{B}_k$, is a pure-strategy NE of game $\mathcal{G}$, but it is not feasible, i.e., it violates at least one of the constraints in (6b). Further, suppose that m is the index of channel which has the most severe violation, i.e., $m = \arg\max_{i \in \mathcal{N}_a} \left(\sum_{j=1}^{N_s} \beta_j^i - 1 \right)$. Thus, $\sum_{j=1}^{N_s} \beta_j^m > 1$. Considering $k \in \mathcal{N}_s$ such that $\beta_k^m > 0$, from (7), we have

$$u_k(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}) = \sum_{i=1}^{N_a} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right) - \mu_m \left(\sum_{j=1}^{N_s} \beta_j^m - 1\right)$$
$$- \sum_{i=1, i \neq m}^{N_a} \mu_i \Theta \left(\sum_{j=1}^{N_s} \beta_j^i - 1\right) \quad (19)$$

Assuming that $\sum_{j=1}^{N_s} \beta_j^m = 1 + \epsilon$ where $\epsilon \neq 0$, there must exist a channel n such that $\sum_{j=1}^{N_s} \beta_j^n \leq \max(0, 1 - \epsilon)$ since $N_a \geq N_s$. Then, an alternative admissible strategy $\tilde{\boldsymbol{\beta}}_k = \left[\tilde{\beta}_k^i\right]_{i \in \mathcal{N}_a} \in \mathcal{B}_k$ can be constructed such that $\tilde{\beta}_k^i =$
$$\begin{cases} \beta_k^i, & i \neq m, n \\ \beta_k^i - \delta, & i = m \\ \beta_k^i + \delta, & i = n \end{cases}$$ where δ is a non-zero discrete

value (i.e., $\delta \in \{\frac{1}{S}, \frac{2}{S}, \ldots, 1\}$) satisfying $\delta \leq \min(1, \epsilon)$, $0 \leq \beta_k^m - \delta \leq 1$ and $0 \leq \beta_k^n + \delta \leq 1$. Note that $\sum_{j=1}^{N_s} \tilde{\beta}_j^n - 1 \leq 1$ since $\delta \leq \min(1, \epsilon)$, and hence, $\Theta\left(\sum_{j=1}^{N_s} \tilde{\beta}_j^n - 1\right) = \Theta\left(\sum_{j=1}^{N_s} \beta_j^n - 1\right)$. Then, $u_k\left(\tilde{\boldsymbol{\beta}}_k, \boldsymbol{\beta}_{-k}\right)$ can be obtained as

$$
\begin{aligned}
u_k(\tilde{\boldsymbol{\beta}}_k, \boldsymbol{\beta}_{-k}) &= \sum_{i=1}^{N_a} \tilde{\beta}_k^i C_k^i (1 - \tilde{\beta}_k^i \alpha_i) - \sum_{i=1}^{N_a} \mu_i \Theta\left(\sum_{j=1}^{N_s} \tilde{\beta}_j^i - 1\right) \\
&= \sum_{i=1}^{N_a} \tilde{\beta}_k^i C_k^i (1 - \tilde{\beta}_k^i \alpha_i) - \mu_m \Theta\left(\sum_{j=1}^{N_s} \beta_j^m - 1 - \delta\right) \\
&\quad - \sum_{i=1, i \neq m}^{N_a} \mu_i \Theta\left(\sum_{j=1}^{N_s} \beta_j^i - 1\right)
\end{aligned}
\tag{20}
$$

Thus, from (19) and (20),

$$
\begin{aligned}
u_k(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}) - u_k(\tilde{\boldsymbol{\beta}}_k, \boldsymbol{\beta}_{-k}) &= \sum_{i=1}^{N_a} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right) - \\
\sum_{i=1}^{N_a} \tilde{\beta}_k^i C_k^i (1 - \tilde{\beta}_k^i \alpha_i) &- \mu_m \left(\sum_{j=1}^{N_s} \beta_j^m - 1\right) \\
+ \mu_m \Theta\left(\sum_{j=1}^{N_s} \beta_j^m - 1 - \delta\right)&
\end{aligned}
\tag{21}
$$

Since $\delta \leq \epsilon$, we have

$$
\begin{aligned}
\mu_m \Theta\left(\sum_{j=1}^{N_s} \beta_j^m - 1 - \delta\right) &- \mu_m \left(\sum_{j=1}^{N_s} \beta_j^m - 1\right) \\
&= \mu_m \Theta(\epsilon - \delta) - \mu_m \epsilon \leq -\mu_m \delta
\end{aligned}
\tag{22}
$$

Subsequently, from (21) and (22),

$$
\begin{aligned}
u_k(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}) - u_k(\tilde{\boldsymbol{\beta}}_k, \boldsymbol{\beta}_{-k}) &\leq \\
\sum_{i=1}^{N_a} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right) - \sum_{i=1}^{N_a} \tilde{\beta}_k^i C_k^i (1 - \tilde{\beta}_k^i \alpha_i) &- \mu_m \delta
\end{aligned}
\tag{23}
$$

Under assumption of $\mu_i > \mu_{th}, \forall i \in \mathcal{N}_a$, from (14), we have

$$
\mu_m > S \sum_{i=1}^{N_a} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right)
\tag{24}
$$

Since $\delta \geq \frac{1}{S}$, from (24), it can be concluded that

$$
\sum_{i=1}^{N_a} \beta_k^i C_k^i \left(1 - \beta_k^i \alpha_i\right) - \mu_m \delta < 0
\tag{25}
$$

Thus, based on (23) and (25),

$$
u_k(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}) - u_k(\tilde{\boldsymbol{\beta}}_k, \boldsymbol{\beta}_{-k}) \leq 0
\tag{26}
$$

This contradicts the assumption that $\boldsymbol{\beta} = \{\boldsymbol{\beta}_k\}_{k=1}^{N_s}$ is a pure-strategy NE of game $\mathcal{G}$ according to the definition of NE in (12). Thus, $\boldsymbol{\beta}$ is not a pure-strategy NE of $\mathcal{G}$. Hence, it can be concluded that all pure-strategy NE must be feasible if $\mu_i > \mu_{th}, \forall i \in \mathcal{N}_a$. ∎

B. Proof of Theorem 4

Let assume that $\boldsymbol{\beta} = \{\boldsymbol{\beta}_k\}_{k=1}^{N_s}$ is the optimal solution of (6). Assuming $\mu_i > \mu_{th}, \forall i \in \mathcal{N}_a$, $\boldsymbol{\beta}$ is the maximizer of the potential function Φ. Based on Theorem 2 in [29], the maximizer of the potential function is the NE of the potential game. Hence, $\boldsymbol{\beta}$ is the NE of the game $\mathcal{G}$.

Subsequently, we need to establish that $\boldsymbol{\beta}$ is the Pareto-optimal NE. Let assume that $\boldsymbol{\beta}$ is not Pareto optimal, and then, there exists an arbitrary strategy profile $\boldsymbol{\beta}' = \{\boldsymbol{\beta}_k'\}_{k=1}^{N_s}$ such that

$$
u_k(\boldsymbol{\beta}_k', \boldsymbol{\beta}_{-k}') \geq u_k(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}), \forall k \in \mathcal{N}_s, k \neq j
\tag{27}
$$

and, for some j,

$$
u_j(\boldsymbol{\beta}_j', \boldsymbol{\beta}_{-j}') > u_j(\boldsymbol{\beta}_j, \boldsymbol{\beta}_{-j})
\tag{28}
$$

As a result,

$$
\sum_{k=1}^{N_s} u_k(\boldsymbol{\beta}') > \sum_{k=1}^{N_s} u_k(\boldsymbol{\beta})
\tag{29}
$$

Since $\boldsymbol{\beta}$ is a NE of the game $\mathcal{G}$, it is feasible based on Theorem 3 (i.e., $\sum_{i=1}^{N_a} \mu_i \Theta\left(\sum_{j=1}^{N_s} \beta_j^i - 1\right) = 0$). Then,

$$
\Phi(\boldsymbol{\beta}) = \sum_{k=1}^{N_s} u_k(\boldsymbol{\beta})
\tag{30}
$$

Furthermore, considering that μ_i's are positive scalars, from (7) and (10), for an arbitrary $\boldsymbol{\beta}'$, we have

$$
\Phi(\boldsymbol{\beta}') \geq \sum_{k=1}^{N_s} u_k(\boldsymbol{\beta}')
\tag{31}
$$

Consequently, based on (29), (30) and (31),

$$
\Phi(\boldsymbol{\beta}') > \Phi(\boldsymbol{\beta})
\tag{32}
$$

This contradicts the fact that $\boldsymbol{\beta}$ is the maximzer of the potential function Φ. Thus, the optimal solution of (6) is the Pareto-optimal pure-strategy NE of the game $\mathcal{G}$. ∎

REFERENCES

[1] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks (Elsevier)*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.

[2] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access: Processing, networking, and regulatory policy," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[3] B. Jabbari, R. Pickholtz, and M. Norton, "Dynamic spectrum access and management," *IEEE Wireless Commun. Mag.*, vol. 17, no. 4, pp. 6–15, Aug. 2010.

[4] L. Berlemann and S. Mangold, *Cognitive Radio and Dynamic Spectrum Access*. Wiley, 2009.

[5] 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.

[6] L. Lai, H. E. Gamal, H. Jiang, and H. Poor, "Cognitive medium access: Exploration, exploitation and competition," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 239–253, Feb. 2011.

[7] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Opportunistic spectrum access via periodic channel sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 785–796, Feb. 2008.

[8] R. Kennedy and P. Ecclesine, "IEEE P802.11af tutorial," IEEE 802.11-10/0742r0, https://mentor.ieee.org/802.11/dcn/10/11-10-0742-00-0000-p802-11af-tutorial.ppt, Jul. 2010.

[9] K. G. Shin, H. Kim, A. W. Min, and A. Kumar, "Cognitive radios for dynamic spectrum access: From concept to reality," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 64–74, Dec. 2010.

[10] C. Ghosh, S. Roy, and D. Cavalcanti, "Coexistence challenges for heterogeneous cognitive wireless networks in TV white spaces," *IEEE Wireless Commun. Mag.*, vol. 18, no. 4, pp. 22–31, Aug. 2011.

[11] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. Netw.*, vol. 8, no. 6, pp. 785–799, Dec. 2000.

[12] C. S. Hwang, K. Seong, and J. M. Cioffi, "Throughput maximization by utilizing multiuser diversity in slow fading random access channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2526–2535, Jul. 2008.

[13] C. S. Hwang and J. M. Cioffi, "Opportunistic CSMA/CA for achieving multi-user diversity in wireless LAN," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2972–2982, Jun. 2009.

[14] H. Kwon, S. Kim, and B. Lee, "Opportunistic multi-channel CSMA protocol for OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1552–1557, May 2010.

[15] M. Derakhshani and T. Le-Ngoc, "Learning-based opportunistic spectrum access with adaptive hopping transmission strategy," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3957–3967, Nov. 2012.

[16] ——, "Opportunistic spectrum access with hopping transmission strategy: A game theoretic approach," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, Quebec City, QC, Canada, Sep. 2012.

[17] C. Alos-Ferrer and N. Netzer, "The logit-response dynamics," *Games and Economic Behavior*, vol. 68, no. 2, pp. 413–427, Mar. 2010.

[18] J. Marden and J. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and a payoff-based implementation," *Games and Economic Behavior*, vol. 75, no. 2, pp. 788–808, Jul. 2010.

[19] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multichannel opportunistic access: Structure, optimality, and performance," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5431–5440, Dec. 2008.

[20] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2795–2808, May 2010.

[21] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.

[22] K. W. Choi and E. Hossain, "Opportunistic access to spectrum holes between packet bursts: A learning-based approach," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2497–2509, Aug. 2011.

[23] L. Rose, S. Lasaulce, S. M. Perlaza, and M. Debbah, "Learning equilibria with partial information in decentralized wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 136–142, Aug. 2011.

[24] P. Mertikopoulos, A. L. Moustakas, and N. Dimitriou, "Vertical handover between wireless service providers," in *Proc. IEEE Intl. symp. on Modeling and Optimization Mobile, Ad Hoc, and Wireless Networks and Workshops (WiOPT)*, Berlin, Germany, Apr. 2008.

[25] H. Zheng and C. Peng, "Collaboration and fairness in opportunistic spectrum access," in *Proc. IEEE Intl. Conf. Commun. (ICC)*, Seoul, Korea, May 2005.

[26] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.

[27] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in WLAN channels: Empirical model and its stochastic analysis," in *Proc. Int. Workshop on Technology and Policy for Accessing Spectrum (TAPAS)*, New York, NY, USA, Aug. 2006.

[28] M. Derakhshani and T. Le-Ngoc, "Intelligent CSMA-based opportunistic spectrum access: Competition and cooperation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012.

[29] D. Monderer and L. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[30] A. B. MacKenzie, L. Dasilva, and W. Tranter, *Game Theory for Wireless Engineers*. Morgan and Claypool Publishers, 2006.

[31] G. Scutari, S. Barbarossa, and D. P. Palomar, "Potential games: A framework for vector power control problems with coupled constraints," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Toulouse, France, May 2006.

[32] D. Okada and O. Tercieux, "Log-linear dynamics and local potential," *Journal of Economic Theory*, vol. 147, no. 3, pp. 1140–1164, May 2012.

[33] S. Lasaulce and H. Tembine, *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*. Academic Press, 2011.

[34] R. Menon, A. B. MacKenzie, J. Hicks, R. M. Buehrer, and J. H. Reed, "A game-theoretic framework for interference avoidance," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1087–1098, Apr. 2009.

Mahsa Derakhshani is a Postdoctoral Research Fellow at the Department of Electrical and Computer Engineering, University of Toronto and also a Research Assistant at the Department of Electrical and Computer Engineering, McGill University. She obtained her Ph.D. degree in Electrical Engineering from McGill University, Montreal, Quebec, Canada in 2013. Prior to that, she received the B.Sc. degree and M.Sc. degree in Electrical Engineering from Sharif University of Technology, Tehran, Iran, in 2006 and 2008, respectively. Since 2014, Dr. Derakhshani has been a recipient of the Fonds de recherche du Québec - Nature et technologies (FQRNT) Postdoctoral Fellowship. Moreover, she was a recipient of the John Bonsall Porter Prize and McGill Engineering Doctoral Award (MEDA) at McGill University from 2008 to 2011. Her research interests include radio resource allocation for wireless networks, applications of convex optimization and game theory for communication systems, and spectrum sensing techniques in cognitive radio networks.
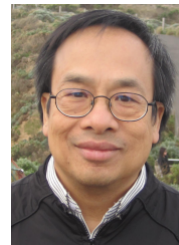
Tho Le-Ngoc obtained his B.Eng. (with Distinction) in Electrical Engineering in 1976, his M.Eng. in 1978 from McGill University, Montreal, and his Ph.D. in Digital Communications in 1983 from the University of Ottawa, Canada. During 1977-1982, he was with Spar Aerospace Limited and involved in the development and design of satellite communications systems. During 1982-1985, he was an Engineering Manager of the Radio Group in the Department of Development Engineering of SRTelecom Inc., where he developed the new point-to-multipoint DA-TDMA/TDM Subscriber Radio System SR500. During 1985-2000, he was a Professor at the Department of Electrical and Computer Engineering of Concordia University. Since 2000, he has been with the Department of Electrical and Computer Engineering of McGill University. His research interest is in the area of broadband digital communications. He is a fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Engineering Institute of Canada (EIC), the Canadian Academy of Engineering (CAE) and the Royal Society of Canada (RSC). He is the recipient of the 2004 Canadian Award in Telecommunications Research, and recipient of the IEEE Canada Fessenden Award 2005. He holds a Canada Research Chair (Tier I) on Broadband Access Communications, and a Bell Canada/NSERC Industrial Research Chair on Performance & Resource Management in Broadband xDSL Access Networks.