Distributed Spectrum Trading in Multiple-Seller Cognitive Radio Networks

Li-Chuan Tseng*, Feng-Tsun Chien*, Ronald Y. Chang†, and Wei-Ho Chung†
*Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan
†Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

***Abstract*—This paper studies spectrum trading in cognitive radio networks in which multiple service providers (SPs) sell unused spectrum to multiple unlicensed secondary users (SUs). Motivated by the nature of the problem with new considerations, spectrum trading is modeled as a multi-leader multi-follower expected Stackelberg game with two levels of competition. The SPs as leaders compete in offering subscription prices (upper-level subgame) and the SUs as followers compete in selecting service from the SPs (lower-level subgame). The lower-level subgame incorporates the time-varying spectrum availability as the external state so that the proposed scheme does not require knowledge of dynamic spectrum availability. To achieve self-organized network operation, we propose decentralized, stochastic learning-based algorithms for the game. The convergence properties of the proposed algorithms toward the Nash equilibrium (NE) are theoretically and numerically studied. The proposed scheme demonstrates good utility performance for the SUs as compared to other service selection schemes.**

I. Introduction

Cognitive radio network (CRN) [1] has been considered as a promising solution to the underutilization of the licensed spectrum. In CRNs, owners of the licensed spectrum are referred to as service providers (SPs). To improve spectrum utilization, the SP may allow secondary users (SUs) to access its licensed spectrum, a behavior called *spectrum trading*. This paper studies spectrum trading from a game-theoretic perspective.

Game theory [2] models the interaction of distributed players and has been an effective tool for studying resource management problems in distributed networks such as CRNs [3] and heterogeneous networks [4]. A game-theoretic approach to spectrum trading was proposed in [5]. The methods in [3]–[5] are suitable for static resource allocation for primary users (PUs), but not for PUs with time-varying behaviors. Specifically, the considered scenario therein was either 1) the quality of service (QoS) requirements of PUs are not flexible and the number of residual channels is fixed, or 2) the QoS requirements of PUs are flexible but the preferences of PUs are fixed. Thus, when the PUs' traffic demands and the channel availability change over time, frequent re-execution of the spectrum trading procedure is needed, resulting in significant overhead in practical CRN operations.

It is therefore motivated to develop efficient spectrum trading strategies for scenarios with time-varying channel/spectrum availability in the game-theoretic context. The considered spectrum trading problem carries a leader-follower structure where the SPs as leaders offer their subscription prices and then the SUs as followers perform service selection according to the offered prices. This hierarchical feature suggests the formulation of a Stackelberg game [2] with two levels of competition. The upper-level subgame corresponds to the price competition among the SPs and the lower-level subgame corresponds to the service selection among the SUs. We aim at developing proper service selection (for SUs) and spectrum pricing (for SPs) strategies by incorporating several practical considerations: 1) players at the same level are unaware of one another and there is no information exchange among them, and 2) the PUs' traffic demands are unknown in the decision-making stage. Due to these considerations, the subgame perfect Nash equilibrium (SPNE) of the Stackelberg game must be defined in an expected manner and cannot be achieved through the traditional backward induction method [2]. In this paper, we propose an *expected* Stackelberg game formulation and develop *fully distributed* selection strategies at both levels which do not require unknown dynamic channel availability. The convergence properties of the proposed algorithm are theoretically and numerically examined.

This work was supported in part by the National Science Council, Taiwan, under grants NSC 102-2218-E-001-001, 102-2221-E-001-006-MY2, 102-2221-E-002-075-MY2, and 103-3113-E-110-002.

II. System Model

We consider a CRN with M SPs and N SUs, where SP_m owns K_m channels in total. The sets of SPs and SUs are denoted by $\mathcal{M}$ and $\mathcal{N}$, respectively. Fig. 1 presents an exemplary CRN with two coexisting SPs. We adopt a shared access model in which the SPs set and announce the subscription prices. An SU sends a request message and pays the price if the SU intends to buy the spectrum opportunities from an SP. SUs have the freedom to dynamically select the SP that will provide the best reward determined by multiple factors (e.g., bandwidth, delay, and price). On the other hand, the SPs adjust their pricing strategies iteratively to improve their own revenues.

When complete information of other SPs is available, the SPs may determine their pricing strategies by anticipating the service selections of SUs (i.e., through backward induction [6], [7]). In our scenario, however, the SPs are independent decision makers who can only learn the pricing strategy through their own iterative updates. The reward is collected when the behaviors of SUs converge, as in [5]. Therefore, the strategy update interval of SPs is longer than that of the SUs. The number of iterations that the SPs wait for the SUs' strategies to converge is denoted as T_{conv}.

We model the spectrum trading as a two-level Stackelberg game. The upper-level subgame corresponds to the competition among SPs in selling the residual channels to the SUs. The lower-level subgame corresponds to the competition among

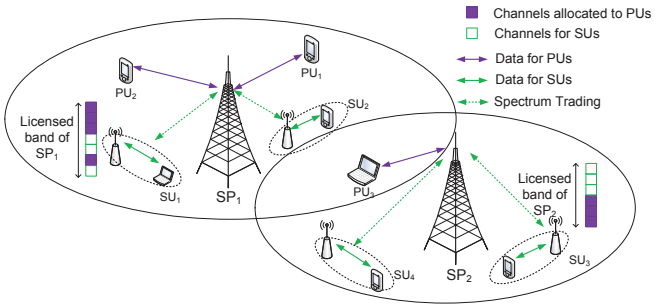

Fig. 1. An exemplary CRN with 2 SPs, 3 PUs, and 4 SUs. The filled and blank blocks in the licensed band of each SP denote the occupied channels by PUs and the residual channels available for SUs, respectively.

SUs in selecting the service from the SPs. To reflect a practical distributed CRN, our system model incorporates the following considerations: 1) Each SU can buy and access the spectrum opportunity of only one SP at a given time. 2) Service selection is done by each SU independently and simultaneously. There is neither negotiation nor sequential updates among SUs. 3) The statistics of the spectrum opportunities are identical among SPs but unknown to the SUs. 4) The number of SUs in the network is unknown to any SP and SU, and the number of SPs is unknown to any SP.

III. Service Selection of Secondary Users

In this section, we present the game-theoretic formulation and self-organized learning procedure in the lower-level subgame. Our objective is to devise for the SUs a fully distributed strategy that takes into account the effect of congestion, offered spectrum opportunities, and the subscription price.

A. Game Model

We model the service selection problem as a noncooperative game where the SUs are the players, and the number of residual channels (after the resource allocation of PUs) is considered as the external state. The game is represented as

$$\mathcal{G}_1 = \left(\mathcal{C}, \mathcal{N}, \{\mathcal{A}_i\}_{i\in\mathcal{N}}, \{u_i\}_{i\in\mathcal{N}}\right)$$

where $\mathcal{C}$ is the space of external states, $\mathcal{N}$ is the set of players, $\{\mathcal{A}_i\}_{i\in\mathcal{N}}$ is the set of actions (service selection) that player i can take, and $\{u_i\}_{i\in\mathcal{N}}$ is the utility of player i. In the service selection game, the utility is defined as the *expected reward* over the external state.

The payoff function is designed to quantify satisfaction levels of SUs, and its value depends on the number of residual channels as well as the number of SUs sharing the same SP. In this paper, we assume the SUs are of the same priority class, and thus the residual channels are equally divided (can be in both frequency and time domains) among them. Let $c_m(j) \in \{0, 1, \ldots, K_m\}$ be the number of residual channels of SP_m at time j, and q_m be the spectrum subscription price paid by each SU that is associated with SP_m (i.e., purchasing spectrum opportunities from SP_m). The price q_m is assumed to take possible values on a pre-defined and finite pricing strategy set of SP_m. Let B_m be the channel bandwidth of SP_m, $\mathcal{N}_m(j)$ be the set of SUs associated with SP_m at time j, and $n_m(j) \triangleq |\mathcal{N}_m(j)|$ is referred to as the *user load* of (i.e., the number of SUs associated with) SP_m at time j. The bandwidth allocated to an SU associated with SP_m at time j is given by $B_m c_m(j)/n_m(j)$, and the instantaneous reward received by SU_i can be given as

$$r_i(j) = \kappa B_m c_m(j)/n_m(j) - q_m, \quad \forall i \in \mathcal{N}_m \tag{1}$$

where constant κ represents the monetary value of unit bandwidth seen by an SU. Without loss of generality, we set $\kappa = 1$ in this paper. The reward function in (1) captures the dynamics of the joint behaviors of multiple SPs and SUs. For notational brevity, we hereafter discard the timing dependence (j) in occasions without ambiguity. With the reward function in (1), the utility (i.e., expected reward) becomes

$$u_i(a_i, a_{-i}) \triangleq \mathbb{E}_{c_{a_i}}\left[r_i|(a_i, a_{-i})\right] = B_{a_i}\bar{c}_{a_i}/n_{a_i} - q_{a_i} \tag{2}$$

where $\bar{c}_{a_i} \triangleq \mathbb{E}[c_{a_i}]$ is the expected number of residual channels of SP_{a_i}. The utility of player i depends on the action of player i (a_i) and of other players (a_{-i}).

B. Analysis of Nash Equilibrium

We assume that the SUs are self-interested and rational players with the objective of maximizing their individual utility, which can be formally described as

$$(\mathcal{G}_1): \max_{a_i\in\mathcal{A}_i} u_i(a_i, a_{-i}), \quad \forall i \in \mathcal{N}. \tag{3}$$

With the utility function in (2), we show the existence of a pure-strategy NE point for the lower-level subgame.

Proposition 1. *The game $\mathcal{G}_1$ is an exact potential game (EPG).*

Proof: Define the function $\Phi : \times_{i\in\mathcal{N}}\mathcal{A}_i \to \mathbb{R}_+$ as

$$\Phi(\mathbf{a}) = \sum_{m=1}^{M}\left(\sum_{l=1}^{n_m}\nu_m(l) - n_m q_m\right) \tag{4}$$

where $\nu_m(l) = B_m\bar{c}_m/l$ and $(n_1, \ldots, n_M)$ is the user load profile resulted from action $\mathbf{a}$. Now, consider that player i changes its action unilaterally from a_i to $\breve{a}_i$. Let n_{a_i} and $n_{\breve{a}_i}$ be the load of SP_{a_i} and $\text{SP}_{\breve{a}_i}$ *before* the change, respectively. Note that player i's change merely affects the SUs subscribing to SP_{a_i} and $\text{SP}_{\breve{a}_i}$, and the change in $\Phi(\cdot)$ due to its unilateral deviation is given by

$$\begin{aligned}
&\Phi(\breve{a}_i, a_{-i}) - \Phi(a_i, a_{-i}) \\
&= \sum_{l=1}^{n_{\breve{a}_i}+1}\nu_{\breve{a}_i}(l) - (n_{\breve{a}_i}+1)q_{\breve{a}_i} + \sum_{l=1}^{n_{a_i}-1}\nu_{a_i}(l) - (n_{a_i}-1)q_{a_i} \\
&- \left(\sum_{l=1}^{n_{\breve{a}_i}}\nu_{\breve{a}_i}(l) - n_{\breve{a}_i}q_{\breve{a}_i} + \sum_{l=1}^{n_{a_i}}\nu_{a_i}(l) - n_{a_i}q_{a_i}\right) \\
&= (\nu_{\breve{a}_i}(n_{\breve{a}_i}+1) - q_{\breve{a}_i}) - (\nu_{a_i}(n_{a_i}) - q_{a_i}) \\
&= u_i(\breve{a}_i, a_{-i}) - u_i(a_i, a_{-i}).
\end{aligned} \tag{5}$$

That is, the changes in $u_i(\cdot)$ and $\Phi(\cdot)$ due to player i's unilateral deviation are identical. Therefore, $\mathcal{G}_1$ is an EPG with potential function $\Phi(\cdot)$. ■

For EPGs, the existence of a pure-strategy NE is always guaranteed and the NE points coincide with the local maximum of the potential function [8].

C. Stochastic Learning Procedure for Service Selection

Here, we propose a decentralized algorithm by which the SUs learn toward the NE strategy profile from their individual action-reward history. The algorithm is based on stochastic learning (SL) [9]. To facilitate the development of the SL-based algorithm, let the mixed strategy $\mathbf{p}_i(j) = \left[p_{i,1}(j), \ldots, p_{i,M}(j)\right]^T$ be the service selection probability vector for player i, where $p_{i,s_i}(j)$ is the probability that player i selects strategy $s_i \in \mathcal{A}_i$ at time j.

A stochastic learning algorithm is characterized by its rule of updating the mixed strategies (based on the action-reward observation), which is usually referred to as the *learning rule*. The proposed self-organized service selection (SoSS) algorithm is described in Algorithm 1. In Algorithm 1, the instantaneous reward serves as a reinforcement signal so that a higher reward leads to a higher probability in the next play (Step 4). Note that the proposed learning algorithm is fully distributed: the selection of SP is based solely on the individual action-reward history without knowledge of other players' actions.

Algorithm 1 Self-organized Service Selection (SoSS)

1: Initially, set $j = 0$, and the spectrum request probability vector as $p_{i,s_i}(j) = 1/|\mathcal{A}_i|, \forall i \in \mathcal{N}, s_i \in \mathcal{A}_i$.
2: At every time instant j, each SU selects an action (i.e., SP) $a_i(j)$ according to $\mathbf{p}_i(j)$.
3: The SUs receive the instantaneous reward specified by (1).
4: Each SU updates its service selection probability vectors according to the following rule:

$$p_{i,s_i}(j+1) = p_{i,s_i}(j) + b \cdot \tilde{r}_i(j)(\mathbb{1}_{\{s_i = a_i(j)\}} - p_{i,s_i}(j)) \quad (6)$$

where $0 < b < 1$ is the learning rate, $\tilde{r}_i(j) = (r_i(j) - r_{\text{inf}})/(r_{\text{sup}} - r_{\text{inf}})$ is the normalized reward with $r_{\text{sup}} = \max_m(B_m K_m - q_m)$ and $r_{\text{inf}} = -q^{\max}$ being respectively the upper and lower bound of the reward, and $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

Proposition 2. *The SoSS Algorithm converges weakly to a (possibly mixed-strategy) NE point when the learning rate b is sufficiently small.*

Proof: This follows from the fact that the SLA converges weakly to a (possibly mixed-strategy) NE point when applied to an ordinal potential game (OPG) [10], and the fact that EPG belongs to OPG. ■

IV. Price Competition among Service Providers

In this section, we present the game model and the fully distributed learning in the upper-level subgame.

A. Game Model

The upper-level price competition game is modeled as a game played by the SPs. The game is represented as a 3-tuple:

$$\mathcal{G}_2 = \left(\mathcal{M}, \{\mathcal{A}_m\}_{m\in\mathcal{M}}, \{u_m\}_{m\in\mathcal{M}}\right)$$

where $\mathcal{M}$ is the set of players (SPs), $\{\mathcal{A}_m\}_{m\in\mathcal{M}}$ is the set of actions (candidate price levels) that player $m, m \in \mathcal{M}$ can take, and $\{u_m\}_{m\in\mathcal{M}}$ is the utility defined as the revenue of SP_m. The revenue that an SP receives depends on its user load at the equilibrium. Given the pricing vector $\mathbf{q} = [q_1, \ldots, q_M]$ and letting $n_m^*(\mathbf{q})$ be the number of SUs associated with SP_m under lower-level NE and q_{-m} be the price of SPs except for SP_m, the utility of SP_m is given by

$$u_m(q_m, q_{-m}) = q_m \cdot n_m^*(\mathbf{q}). \quad (7)$$

As in the lower-level game, we adopt learning algorithm for the SPs to adapt to proper pricing strategies. In the upper-level competition, as can be seen from (7), the revenues received by the SPs depend on the user loads at the NE point of the lower-level game. If the user loads (and therefore the revenues) vary (because of non-uniqueness of NE) under a fixed pricing strategy profile, it could be difficult for the SPs to evaluate each strategy and find proper ones.

Fortunately, while there may be multiple NE points, the user loads at NE are unique with a fixed subscription price vector when the number of SUs is very large. Following the discussions in [5], [7], it is known that for a given subscription price vector, the unique steady-state user loads are characterized by the Wardrop equilibrium [11]. At the equilibrium, the per-SU utilities provided by subscribing to the SPs that have one or more associated SUs are equal, and are greater than that of an SU associated with an unused SP. In other words, $\forall m, m' \in \mathcal{M}$ with $n_m^* > 0$, we have

$$\begin{cases} \nu_m(n_m^*) - q_m = \nu_{m'}(n_{m'}^*) - q_{m'}, & \text{if } n_{m'}^* > 0, \\ \nu_m(n_m^*) - q_m > \nu_{m'}(1) - q_{m'}, & \text{otherwise.} \end{cases} \quad (8)$$

Remark 1. The unique user load profile under Wardrop equilibrium is based on the assumption that the game is non-atomic, i.e., the number of SUs is big compared to the number of SPs. When the number of SUs is finite, solving (8) may result in non-integer n_m^*'s, which do not constitute feasible user loads, and the user load profiles under lower-level NE may be different. However, the user load profile under NE is unique when an additional constraint is applied.

Definition 1 (Strict Nash Equilibrium)**.** An action profile $\mathbf{a}^* = (a_1^*, \ldots, a_N^*)$ is a pure-strategy strict NE of the non-cooperative game $\mathcal{G}$ if and only if unilateral deviation will result in decreased utility for all players, i.e.,

$$u_i(a_i^*, a_{-i}^*) > u_i(a_i, a_{-i}^*), \quad \forall i \in \mathcal{N}, \forall a_i \in \mathcal{A}_i \setminus \{a_i^*\}. \quad (9)$$

Proposition 3. *The user load profile is unique for all NE points if the lower-level NE is strict.*

Proof: We show the proposition by using an important property of potential games: an NE always coincides with a local maximum of the potential function.

For a network with M SPs, we let $(\hat{n}_1, \ldots, \hat{n}_M)$ be the user load profile resulted from an NE $\hat{\mathbf{a}}$. Let $\mathbf{a}_1$ denote the action profile that corresponds to an SU unilaterally deviating from the NE $\hat{\mathbf{a}}$ by associating with SP_i rather than with SP_j originally, which yields a new user load profile $(\hat{n}_1, \ldots, \hat{n}_i + 1, \ldots, \hat{n}_j - 1, \ldots, \hat{n}_M)$. Then, from (4) we have

$$\Phi(\mathbf{a}_1) - \Phi(\hat{\mathbf{a}}) = \frac{B_i \bar{c}_i}{\hat{n}_i + 1} - q_i - \frac{B_j \bar{c}_j}{\hat{n}_j} + q_j < 0 \quad (10)$$

where the inequality follows from the fact that the NE is a

Algorithm 2 Self-organized Pricing (SoP)

1: Initially, set $k = 0$. Set the estimated utility and the pricing probability vector as $(\forall m \in \mathcal{M}, s_m \in \mathcal{A}_m)$

$$\hat{u}_{m,s_m}(-1) = 0 \quad \text{and} \quad p_{m,s_m}(0) = 1/|\mathcal{A}_m|.$$

2: At the beginning of the kth iteration, each seller selects an action $a_m(k)$ according to the current pricing strategy $\mathbf{p}_m(k)$.

3: When the service selection of SUs converges, each seller m receives the utility $u_m(k)$ specified by (2) depending on the user load.

4: All SPs update their estimated utility and pricing probability vector in iteration k according to the following rules:

$$\begin{aligned} &\hat{u}_{m,s_m}(k) - \hat{u}_{m,s_m}(k-1) \\ &\quad = \eta \mathbb{1}_{\{a_m(k)=s_m\}} \left(u_m(k) - \hat{u}_{m,s_m}(k-1)\right) \\ &p_{m,s_m}(k+1) = \frac{p_{m,s_m}(k)(1+\epsilon)^{\hat{u}_{m,s_m}(k)}}{\sum_{s'_m \in \mathcal{A}_m} p_{m,s'_m}(k)(1+\epsilon)^{\hat{u}_{m,s'_m}(k)}} \end{aligned} \tag{12}$$

where η and ϵ are the learning rates for utility estimation and pricing probability updating, respectively.

local maximizer of the potential function and is strict here.

Now suppose that there exists another NE $\hat{\mathbf{a}}'$ that results in a different user load profile $(\hat{n}_1 + d_1, \hat{n}_2 + d_2, \ldots, \hat{n}_M + d_M)$ for some integer $d_i \neq 0$, $i = 1, \ldots, M$. Note that we have the constraint $\sum_{m=1}^{M} d_m = 0$ here to maintain the same total number of SUs. With the constraint, there must exist at least one $d_i \geq 1$ and one $d_j \leq -1$ for some i and j. Consider the action profile $\mathbf{a}'_1$ that corresponds to an SU unilaterally deviating from the NE $\hat{\mathbf{a}}'$ by associating with SP_j rather than with SP_i originally. This action $\mathbf{a}'_1$ yields a new user load profile $(\hat{n}_1+d_1, \ldots, \hat{n}_i+d_i-1, \ldots, \hat{n}_j+d_j+1, \ldots, \hat{n}_M+d_M)$. Then, the potential difference between actions $\hat{\mathbf{a}}'$ and $\mathbf{a}'_1$ is

$$\Phi(\hat{\mathbf{a}}') - \Phi(\mathbf{a}'_1) = \frac{B_i \bar{c}_i}{\hat{n}_i + d_i} - q_i - \frac{B_j \bar{c}_j}{\hat{n}_j + d_j + 1} + q_j < 0 \tag{11}$$

where the inequality follows from $d_i \geq 1$, $d_j \leq -1$, and the result in (10). Note that (11) contradicts the fact that an NE is always a local maximizer of the potential function. Thus, by contradiction we conclude that the NE $\hat{\mathbf{a}}'$ yields the same user load profile as the NE $\hat{\mathbf{a}}$. ■

B. Stochastic Learning Procedure for Price Competition

In the upper-level subgame, learning-based algorithms help the SPs gradually adjust their pricing strategies based on the service selections of the SUs at the equilibrium of the lower-level subgame. A seller's pricing strategy is defined over a probability space of its candidate price levels.

As in the lower-level subgame, two main issues are considered when designing the learning algorithm for the upper-level subgame, namely, the learning rule and the convergence property. First, since the total number of SUs is unknown to the SPs, it is difficult to obtain the upper bound and normalize the revenue. Therefore, a probability update rule different from (6) is needed. In this paper, we consider the *multiplicative-weight* rule for mixed-strategy update. The learning procedure in the self-organized price (SoP) competition is described in Algorithm 2. The multiplicative-weight update rule in (12) belongs to the combined fully distributed payoff and strategy reinforcement learning (CODIPAS-RL) [12], in which learning applies to both the expected payoff and the strategies.

The second issue is the convergence behavior when the SL algorithm is applied in the price competition game. Unlike the lower-level subgame, the upper-level competition with the utility in (8) is not a potential game. Thus, the convergence toward an NE point in the upper-level subgame is not theoretically verified. We will numerically demonstrate the convergence of the proposed algorithm in Sec. V.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Number of SPs	$M = 2$
Total channel bandwidth of SPs	$B_1 = B_2 = 3$
Max. number of channels	$K_m = 3$
Ch. availability of SP_1	$\mathbf{x}_1 = [0, 0.1, 0.3, 0.6]$
Ch. availability of SP_2	$\mathbf{x}_2 = [0, 0.4, 0.3, 0.3]$
Pricing strategies	$\mathcal{A}_m = [1, 1.5, 2, 2.5], \forall m$
Learning rate	$(\eta, \epsilon) = (0.1, 0.05)$
Number of SUs	$N = 6$
Learning rate of SUs	$b = 0.3$
Waiting time for obtaining NE	$T_{\text{conv}} = 400$

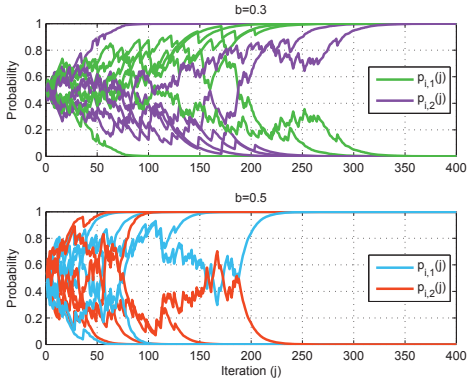

Fig. 2. Evolution of the mixed strategies (probability of taking different actions) of all players. Each pair of $p_{i,1}(j)$ and $p_{i,2}(j)$ shows the behavior of a player $i \in \mathcal{N}$.

V. NUMERICAL RESULTS

Here, we present simulation results to evaluate the performance of the proposed scheme. The distribution of the number of residual channels (i.e., spectrum opportunities) offered by SP_m is described by a vector $\mathbf{x}_m = [x_{m,0}, \ldots, x_{m,c}, \ldots, x_{m,K_m}]$, where $x_{m,c}$ denotes the probability that SP_m possesses c residual channels. The default values of simulation parameters are given in Table I, and these values are adopted in the simulations unless otherwise specified. We first study the lower-level game under a given price vector $(q_1, q_2) = (1, 1.5)$. The purpose is to observe the convergence behavior and the performance of the proposed algorithm. Then, we study the upper-level game.

A. Lower-level Subgame

1) Convergence Behavior: Fig. 2 shows the evolutions of the choice probabilities of the actions (i.e., mixed strategies) for service selection using the proposed SL algorithm. With equal initial probabilities, it is observed that the service selections converge to pure strategies in 350 and 250 iterations

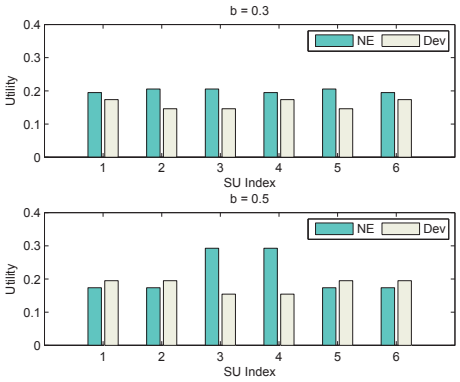

Fig. 3. Test of unilateral deviation from the learned strategy profile of each of the $N = 6$ players, with learning rates $b = 0.3$ and $b = 0.5$.

for $b = 0.3$ and $b = 0.5$, respectively. We observe that the final user loads of SUs are different: $(n_1, n_2) = (3, 3)$ for $b = 0.3$ and $(n_1, n_2) = (4, 2)$ for $b = 0.5$. To verify the NE property of the converged point, we test the unilateral deviation from the learned service selection strategies of each of the $N = 6$ players. The comparison in terms of (normalized) utilities is given in Fig. 3. It is seen that when $b = 0.3$, unilateral deviation from the learned strategy results in a lower utility for all players. This confirms that the outcome of the learning algorithm is an NE point. However, when $b = 0.5$, four SUs (#1, #2, #5, and #6) gain higher utility by unilateral deviation, which implies that the resulting strategy is not an NE point. Combined with the results in Fig. 2, we observe that the user load under NE is $(n_1, n_2) = (3, 3)$. Since the learning algorithm converges to $(n_1, n_2) = (4, 2)$ when $b = 0.5$, any of the SUs subscribing to SP_1 can improve its utility by unilaterally deviating to SP_2.

2) Performance Comparison: We compare the performance of the proposed service selection scheme with two other approaches, namely, random selection and exhaustive search. In the random selection scheme, each SU randomly subscribes to a network in each iteration. Neither learning algorithm nor centralized controller is implemented. Since the SUs possess very little knowledge on the environments, the random selection scheme is an intuitive heuristic leveraging the randomness of the external states. In the exhaustive search scheme, it is assumed that there exists a centralized controller with all the system information including the numbers of SUs and SPs, and the channel availability statistics. The service selection profile is determined by maximizing the expected sum utility, i.e., finding the optimal action profile $\mathbf{a}^{\text{opt}} = (a_1, \ldots, a_N)$ where $\mathbf{a}^{\text{opt}} = \arg\max_{\mathbf{a}} \sum_{i=1}^{N} u_i(a_i, a_{-i})$.

The performance of different service selection schemes is evaluated by the average normalized utility per SU, as shown in Fig. 4. The learning method achieves around 87% of the exhaustive search, and outperforms the random selection by about 20%.

B. Upper-level Subgame

1) Convergence Behavior: For the upper-level subgame, again we first study the convergence behaviors of the pricing

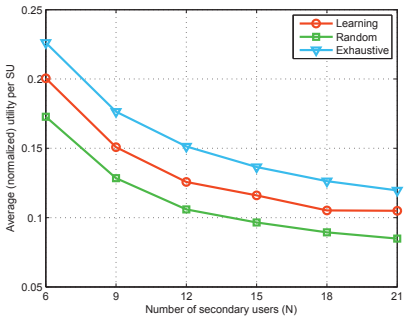

Fig. 4. Comparison of the average (normalized) utility per SU for different service selection schemes.

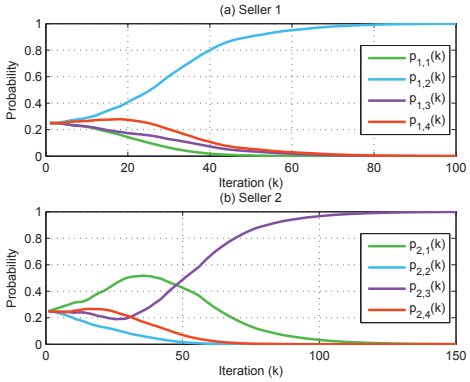

Fig. 5. Evolution of the mixed strategies (probability of taking different actions) of the $M = 2$ sellers.

strategies. Fig. 5 shows the evolution of the choice probabilities of the pricing strategies using Algorithm 2. With equal initial probabilities, it is observed that the pricing probabilities converge to pure strategies in around 80 and 120 cycles for seller 1 and seller 2, respectively. When the competition converges to equilibrium, seller 1 sets subscription price level 2 ($q_1 = 1.5$) and seller 2 sets subscription price level 3 ($q_2 = 2$).

The performance of the learned strategy profile is studied in Fig. 6. For each of the two sellers, four different pricing strategies are examined by fixing the opponent's strategy at the learning result of Algorithm 2. The first performance metric is the individual revenue which can be used for the verification of NE, and the results are provided in Fig. 6(a). It is shown that when SP_2 sticks to the NE strategy (i.e., $q_2 = 1.5$), SP_1 gets the best utility by also taking the NE strategy ($q_2 = 2$). Similarly, SP_2 must follow its NE strategy when SP_1 does so. On the other hand, as the second performance metric, the results of total revenue (defined as the sum of the revenues of the two SPs) are shown in Fig. 6(b). While unilateral deviation from the learned strategy does improve the total revenue in some cases, the seller who benefits (i.e., obtains higher revenue) is the opponent instead of the deviating seller.

Note that the sum revenue (i.e., the total utility in the upper-level competition) is maximized when both sellers set the highest price level (i.e., $q_1 = q_2 = 2.5$). In this case, the maximum sum revenue is 15. The sum revenue obtained by using the NE

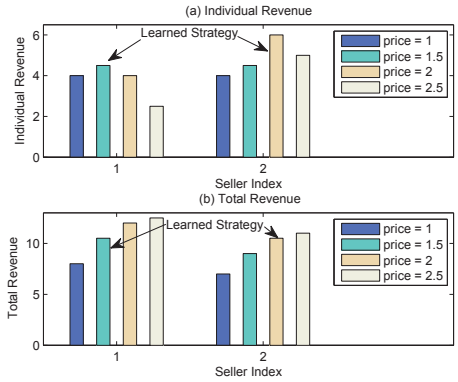

Fig. 6. Test of different strategies. For each seller, the four bars show its revenues when taking the four different pricing strategies, while its opponent sticks to the learned strategy.

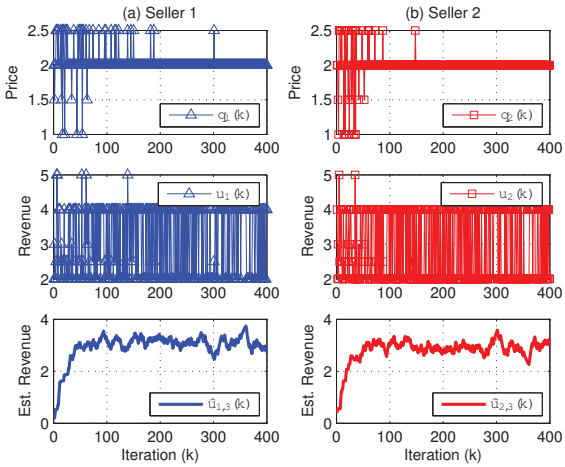

Fig. 7. Effects of nonunique user load profiles at lower-level NEs on the prices, revenues, and estimated revenues for both sellers in the upper-level game.

pricing strategy is 10.5, which is only 70% of the maximum value. From the sellers' point of view, this is an *efficiency loss* as a consequence of the game-theoretic formulation and price competition. Despite the efficiency loss of the SPs, the SUs do pay less when the sellers are competing with each other instead of cooperating. This observation confirms the nature of a free spectrum market: the noncooperative nature of the sellers prevents the collusion among sellers and buyers benefit from the price competition of the sellers.

2) Effects of User Loads: The user load profile resulted from all lower-level NE points is unique in all the simulation results presented so far. However, as mentioned in Sec. IV-A, when the number of SUs is finite, the user load profiles under lower-level NE may be different. To study the effect of user load profiles, we consider a scenario of $N = 3$ SUs and two sellers where each seller only has one residual channel available. Fig. 7 shows the dynamics of prices, revenues, and estimated revenues for both sellers. It is observed that the pricing strategies converge to the subscription price level 3 (i.e., $q_1 = q_2 = 2$), which is an NE point in the upper-level subgame, for both sellers. On the other hand, the revenue (utility) oscillates between $u_m(k) = 2$ and $u_m(k) = 4$ for both sellers, which means that the user loads oscillate between $(n_1, n_2) = (4, 2)$ and $(n_1, n_2) = (2, 4)$. The two user load profiles correspond to two NEs in the lower-level game. The estimated revenues converge to around $\hat{u}_m = 3$ for $m = 1, 2$. If we consider the user loads as an external state and assume that the two user load profiles are equally likely (i.e., $\Pr\{(n_1, n_2) = (4, 2)\} = \Pr\{(n_1, n_2) = (2, 4)\} = 0.5$), the utilities (i.e., expected reward over external states) are $u_m = 3$ for $m = 1, 2$.

VI. Conclusion

In this paper, we have studied the problem of spectrum trading with multiple sellers and time-varying spectrum opportunities. We have formulated spectrum trading as a two-level Stackelberg game whose upper- and lower-level subgames model the price competition of SPs and the service selection of SUs, respectively. Distributed stochastic learning-based algorithms have been proposed for the strategic learning at both levels. Simulation results demonstrated the convergence of the algorithms towards NE at both levels. In the lower-level subgame, the proposed method achieves a comparable average utility performance as the centralized exhaustive search scheme and outperforms the random selection scheme. In the upper-level subgame, despite the reduced sum revenue for the SPs, the price competition among SPs benefits the SUs as the SUs pay less when the SPs are competing with each other.

References

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

[2] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.

[3] B. Wang, Y. Wu, and K. J. R. Liu, "Game theory for cognitive radio networks: An overview," *Computer Netw.*, vol. 54, no. 14, pp. 2537–2561, 2010.

[4] R. Trestian, O. Ormond, and G.-M. Muntean, "Game theory-based network selection: Solutions and challenges," *IEEE Commun. Surv. Tut.*, vol. 14, no. 4, pp. 1212–1231, 2012.

[5] D. Niyato, E. Hossain, and Z. Han, "Dynamics of multiple-seller and multiple-buyer spectrum trading in cognitive radio networks: A game-theoretic modeling approach," *IEEE Trans. Mobile Comput.*, vol. 8, no. 8, pp. 1009–1022, Aug. 2009.

[6] L. Duan, J. Huang, and B. Shou, "Duopoly competition in dynamic spectrum leasing and pricing," *IEEE Trans. Mobile Comput.*, vol. 11, no. 11, pp. 1706–1719, Nov. 2012.

[7] J. Elias, F. Martignon, L. Chen, and E. Altman, "Joint operator pricing and network selection game in cognitive radio networks: Equilibrium, system dynamics and price of anarchy," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4576–4589, Nov. 2013.

[8] D. Monderer and L. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.

[9] P. Sastry, V. Phansalkar, and M. Thathachar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 5, pp. 769–777, May 1994.

[10] L.-C. Tseng, F.-T. Chien, D. Zhang, R. Y. Chang, W.-H. Chung, and C.-Y. Huang, "Network selection in cognitive heterogeneous networks using stochastic learning," *IEEE Commun. Lett.*, vol. 17, no. 12, pp. 2304–2307, Dec. 2013.

[11] J. G. Wardrop, "Some theoretical aspects of road traffic research." in *ICE Proceedings: Engineering Divisions*, vol. 1, no. 3. Thomas Telford, 1952, pp. 325–362.

[12] H. Tembine, *Distributed Strategic Learning for Wireless Engineers*. CRC Press, 2012.