Network Selection in Cognitive Heterogeneous Networks Using Stochastic Learning

Li-Chuan Tseng, Feng-Tsun Chien, Daqiang Zhang, Ronald Y. Chang, Wei-Ho Chung, and ChingYao Huang

***Abstract*—Coexistence of multiple radio access technologies (RATs) is a promising paradigm to improve spectral efficiency. This letter presents a game-theoretic network selection scheme in a cognitive heterogeneous networking environment with time-varying channel availability. We formulate the network selection problem as a noncooperative game with secondary users (SUs) as the players, and show that the game is an ordinal potential game (OPG). A decentralized, stochastic learning-based algorithm is proposed where each SU's strategy progressively evolves toward the Nash equilibrium (NE) based on its own action-reward history, without the need to know actions in other SUs. The convergence properties of the proposed algorithm toward an NE point are theoretically and numerically verified. The proposed algorithm demonstrates good throughput and fairness performances in various network scenarios.**

***Index Terms*—Heterogeneous networks, cognitive radio, self-organized network selection, stochastic learning.**

I. Introduction

THE rapid increase of wireless applications has rendered the single-network wireless system insufficient in meeting the traffic demands due to the inefficient spectrum utilization. A heterogeneous network, where multiple radio access technologies (RATs) coexist, has emerged as a viable alternative solution. In a heterogeneous network, users are allowed to access the spectrum licensed to different spectrum owners, which are called service providers (SPs), and as a result a more efficient spectrum utilization can potentially be achieved. In heterogeneous networks, one significant issue is the network selection where each user determines which network to associate with. Early works on heterogeneous networks primarily focused on the study of vertical handover procedures for mobile devices [1]. The handover decision is made independently at each user according to the user's received signal strength (RSS) from different SPs and the predicted trajectory of movement. However, even a user is associated with an SP of good RSS, its achievable throughput may degrade when the total number of users served by the same SP increases.

To this end, recent works consider the joint behaviors of users in the decision-making process for network selection, particularly from a game-theoretic perspective [2]–[4]. Evolutionary game framework was applied to cognitive heterogeneous networks in [2]. The proposed method therein requires knowledge of other users' actions and thus cannot be implemented in a fully distributed manner. Khan *et al.* [3] proposed a distributed hybrid learning method for 4G heterogeneous networks. The convergence towards a pure strategy profile was demonstrated without indicating whether or not the achieved strategy profile is an equilibrium point. A more extensive survey can be found in the work by Trestian *et al.* [4].

In this letter, we consider the problem of network selection in a heterogeneous network featuring cognitive radio (CR). Specifically, we consider the *primary network access* scenario [5] where both primary users (PUs) and secondary users (SUs) are served by the primary networks. We model the network selection by SUs as a noncooperative game, where the number of residual channels (determined by the time-varying demands of PUs) is considered as the external state. With our proposed utility function, the game is shown to be an ordinal potential game (OPG) [6]. A stochastic learning algorithm (SLA) is proposed to perform network selection independently at each SU based only on its action-reward history; the knowledge of other SUs' actions is not needed. The convergence property of the algorithm to a Nash equilibrium (NE) point is verified theoretically and numerically. To the best of our knowledge, this work presents the first application of SLA to OPGs in wireless networks. Notably, formulating an OPG poses fewer constraints on the design of utility functions as compared with the exact potential game (EPG) [6], and thus facilitates mapping practical resource management problems in distributed networks into proper game-theoretic formulations.

II. System Model

We consider a cognitive heterogeneous network with M SPs and N SUs. The sets of SPs and SUs are denoted by $\mathcal{M}$ and $\mathcal{N}$, respectively. SP$_m$ owns K_m channels. At time instant j, after resource allocation for PUs, SP$_m$ has $C_m(j)$ residual channels that can be used to serve the SUs. Fig. 1 presents an exemplary heterogeneous network where two RATs coexist.

To reflect a practical wireless heterogeneous network, our system model incorporates the following considerations:

1) Due to hardware and protocol limitations, each SU can subscribe to only one SP at a given time.
2) Each SU selects the SP independently. There is neither central control nor negotiation among SUs.
3) The statistics of the number of residual channels owned by each SP are fixed but unknown to the SUs.
4) The number of SUs in the system, N, is unknown.

Manuscript received August 16, 2013. The associate editor coordinating the review of this letter and approving it for publication was A. Rabbachin.

This work was supported in part by the National Science Council, Taiwan, under grants NSC 102-2218-E-001-001, NSC 101-2219-E-009-018, and NSC 102-2220-E-009-042.

L.-C. Tseng, F.-T. Chien, and C. Y. Huang are with the Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan (e-mail: lctseng.ee90@nctu.edu.tw, {ftchien, cyhuang}@mail.nctu.edu.tw).

D. Zhang is with the School of Software Engineering, Tongji University, Shanghai, China (e-mail: dqzhang@ieee.org).

R. Y. Chang and W.-H. Chung are with the Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan (e-mail: {rchang, whc}@citi.sinica.edu.tw).

Digital Object Identifier 10.1109/LCOMM.2013.102113.131876

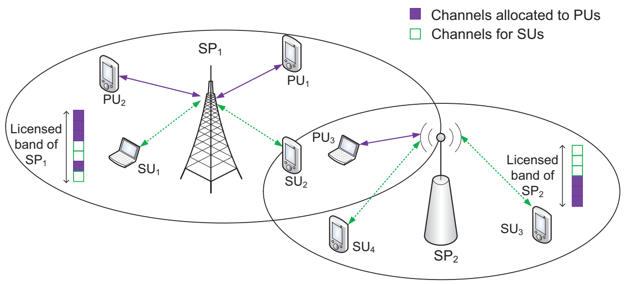

Fig. 1. An exemplary heterogeneous network with 2 SPs, 3 PUs, and 4 SUs. The filled and blank blocks in the licensed band of each SP denote the busy channels currently used by the PUs and the residual channels available for serving the SUs, respectively.

TABLE I
SUMMARY OF SYMBOLS FOR GAME-THEORETIC FORMULATION

Symbol	Meaning
$\mathcal{N}$	the set of SUs
$\mathcal{M}$	the set of SPs
$\mathcal{C}$	external state space (channel availability)
$C_m(j)$	number of available channels of SP_m at time j
$\mathcal{A}_i \subseteq \mathcal{M}$	the set of actions of player i
$s_i \in \mathcal{A}_i$	an element of $\mathcal{A}_i$
$a_i(j) \in \mathcal{A}_i$	the action (SP selection) of player i at time j
$a_{-i}(j) \in \mathcal{A}_i$	actions of players except for i at time j
$\mathcal{P}_i := \Delta(\mathcal{A}_i)$	the set of probability distribution over $\mathcal{A}_i$
$\mathbf{p}_i(j) \in \mathcal{P}_i$	mixed strategy of player i at time j
$r_i(j) \in \mathbb{R}$	instantaneous reward of player i at time j

Notably, the only information available for decision making is the action-reward history of individual players (SUs).

Let $\mathcal{N}_m(j) = \{i \in \mathcal{N} | a_i(j) = m\}$ be the set of SUs associated with SP_m at time j, where $a_i(j)$ is the action (i.e., network selection) of SU_i at time j. Here, we consider the case where the SUs are of the same priority class, and thus the residual channels are equally divided (can be in both frequency and time domain) to them. Then, if $a_i(j) = m$, the throughput of SU_i at time j is given by

$$r_i(j) = C_m(j) R_{m,i} / n_m(j), \quad \forall i \in \mathcal{N}_m \tag{1}$$

where $n_m(j) \triangleq |\mathcal{N}_m(j)|$ and $R_{m,i}$ is the per-channel throughput of SU_i when SU_i is the only user associated with SP_m. The value of $R_{m,i}$ is determined by the modulation order (e.g., $R_{m,i} = 4$ when 16-QAM is adopted). For notational brevity, we hereafter discard the timing dependence in occasions without ambiguity.

III. SELF-ORGANIZED NETWORK SELECTION

In this section, we present the game-theoretic formulation of the self-organized network selection problem. The notations used in the formulation are summarized in Table I.

A. Game Model

We model the network 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:

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

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 (network selection) that player i can take, and $\{u_i\}_{i\in\mathcal{N}}$ is the utility function of player i that depends on the actions of itself as well as other players.

The SUs are selfish and rational players with the objective of maximizing their individual throughput. Thus, we define the instantaneous reward of player i at time j as the throughput specified in (1). The reward function captures the dynamics of the behavior of PUs as well as the joint behaviors of multiple SUs. Then, we define the utility function as the expected reward of player i over the channel availability[1], i.e.,

$$u_i(a_i, a_{-i}) \triangleq \mathbb{E}_{C_{a_i}}\left[r_i | (a_i, a_{-i})\right] = \frac{R_{a_i,i}}{n_{a_i}} \sum_{k=1}^{K_{a_i}} x_{a_i,k} \cdot k \tag{2}$$

where $x_{a_i,k}$ is the probability of $C_{a_i} = k$ with $\sum_{k=1}^{K_{a_i}} x_{a_i,k} = 1$, and n_{a_i} is the number of players taking action a_i, which depends on the action of player i (a_i) as well as other players' actions (a_{-i}). Formally, the game can be expressed as

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

B. Analysis of Nash Equilibrium

With the utility function in (2), we show the existence of an NE point for the considered game here.

Proposition 1. *The game $\mathcal{G}$ is an OPG.*

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

$$\Phi(a_1, \ldots, a_N) = \prod_{m=1}^{M} \prod_{l=1}^{n_m} \nu_m(l) \cdot \prod_{i=1}^{N} R_{a_i,i} \tag{4}$$

where

$$\nu_m(l) \triangleq \frac{1}{l} \sum_{k=1}^{K_m} x_{m,k} \cdot k \tag{5}$$

is the average number of channels allocated by SP_m to each of its SUs when there are l SUs associated with SP_m. 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 number of SUs associated with SP_{a_i} and $\text{SP}_{\breve{a}_i}$ *before* the change, respectively. If this change improves the u_i, from the definitions in (2) and (5), we have

$$\begin{aligned} & u_i(\breve{a}_i, a_{-i}) > u_i(a_i, a_{-i}) \\ \Leftrightarrow & \nu_{\breve{a}_i}(n_{\breve{a}_i} + 1) \cdot R_{\breve{a}_i,i} > \nu_{a_i}(n_{a_i}) \cdot R_{a_i,i}. \end{aligned} \tag{6}$$

Meanwhile, since player i's change merely affects resource allocations in SP_{a_i} and $\text{SP}_{\breve{a}_i}$, the change in Φ caused by player i's unilateral deviation is given by

$$\frac{\Phi(\breve{a}_i, a_{-i})}{\Phi(a_i, a_{-i})} = \frac{\nu_{\breve{a}_i}(n_{\breve{a}_i} + 1) \cdot R_{\breve{a}_i,i}}{\nu_{a_i}(n_{a_i}) \cdot R_{a_i,i}} > 1. \tag{7}$$

From (6) and (7) we find that the variations in u_i and Φ due to player i's unilateral deviation have the same sign, i.e.,

$$u_i(\breve{a}_i, a_{-i}) - u_i(a_i, a_{-i}) > 0 \Leftrightarrow \Phi(\breve{a}_i, a_{-i}) - \Phi(a_i, a_{-i}) > 0. \tag{8}$$

[1]The same formulation can be applied under fading channels, where the time-varying $R_{m,i}$ is considered as part of the external state and its average value is adopted in u_i. A longer learning period may be required in this case.

Therefore, $\mathcal{G}$ is an OPG with potential function Φ [6]. ■

The existence of a pure-strategy NE is always guaranteed and it coincides with a local maximum of the potential function [6]. Note that an EPG formulation [7] requires

$$u_i(\breve{a}_i, a_{-i}) - u_i(a_i, a_{-i}) = \Phi(\breve{a}_i, a_{-i}) - \Phi(a_i, a_{-i}). \quad (9)$$

Comparing (8) and (9), it is observed that the constraint on the utility function is relaxed in OPG, which facilitates game-theoretic developments.

C. Stochastic Learning Procedure

Here, we discuss obtaining the NE via stochastic learning. As the channel availability is time-varying and the action is selected by each player simultaneously and independently in each play, previously developed algorithms requiring complete information (e.g., better response dynamics [6]) may not be applicable. To this end, we propose a decentralized algorithm based on stochastic learning (SL) [8], by which the SUs learn toward the equilibrium strategy profile from their individual action-reward history.

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 network 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. The proposed self-organized network selection (SoNS) algorithm is described in Algorithm 1.

Algorithm 1 Self-organized Network Selection (SoNS)

1: Initially, set $j = 0$, and the network selection 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 j, each player selects an action $a_i(j)$ as the outcome of a probabilistic experiment based on $\mathbf{p}_i(j)$.
3: The SUs receive the instantaneous reward $r_i(j)$ specified by (1) from the SPs.
4: Each SU updates its network selection probability vectors according to the following rule:

$$p_{i,s_i}(j+1) = p_{i,s_i}(j) + b \cdot \tilde{r}_i(j)(\mathbb{1}_{\{s_i = a_i(j)\}} - p_{i,s_i}(j)) \quad (10)$$

where $0 < b < 1$ is the learning rate, $\mathbb{1}_{\{\cdot\}}$ is the indicator function, and $\tilde{r}_i(j)$ is the normalized reward.

The instantaneous reward (throughput) serves as a reinforcement signal so that a high reward brings a high probability in the next strategy update (Step 4). Also note that network selection based on a probabilistic experiment (Step 2) might result in handover between different networks in the beginning of the learning procedure. However, a stable long-term network selection strategy will be yielded after the learning period (Proposition 2) and the time required for convergence is a small fraction of the total operation time.

Proposition 2. *The SoNS Algorithm converges to NE when the learning rate b is sufficiently small.*

Proof: Let $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_N)$ be the mixed strategy profile of all players. Define $\psi_i(\mathbf{P}) \triangleq \mathbb{E}_{\mathbf{P}}[u_i]$ and $\Psi(\mathbf{P}) \triangleq \mathbb{E}_{\mathbf{P}}[\Phi]$ be the expected reward function of player i and the expected potential function, respectively, over the mixed strategy $\mathbf{P}$. Let $\mathbf{e}_{s_i}$ be a unit probability vector (of appropriate dimension) with the s_i-th component being unity and all others zero, and $\mathbf{P}_{-i}$ be the mixed strategy of players except for i. We have $\Psi(\mathbf{P}) = \sum_{s_i} p_{i,s_i} \Psi(\mathbf{e}_{s_i}, \mathbf{P}_{-i})$, and $\partial\Psi(\mathbf{P})/\partial p_{i,s_i} = \Psi(\mathbf{e}_{s_i}, \mathbf{P}_{-i})$. Also, the sequence $p_{i,s_i}(j)$ converges to the solution of the ODE [8, Theorem 3.1]:

$$\frac{dp_{i,s_i}(t)}{dt} = p_{i,s_i}(t) \sum_{s_i'} p_{i,s_i'}(t) \left[\psi_i(\mathbf{e}_{s_i}, \mathbf{P}_{-i}) - \psi_i(\mathbf{e}_{s_i'}, \mathbf{P}_{-i})\right]$$

for all i, s_i, where $p_{i,s_i}(t)$ is the continuous time extension of $p_{i,s_i}(j)$. Consider the derivative of $\Psi(\mathbf{P})$ with respect to t:

$$\begin{aligned}\frac{d\Psi(\mathbf{P})}{dt} &= \sum_i \sum_{s_i} \frac{\partial\Psi(\mathbf{P})}{\partial p_{i,s_i}} \frac{dp_{i,s_i}(t)}{dt} \\ &= \sum_i \sum_{s_i, s_i'} p_{i,s_i}(t) p_{i,s_i'}(t) \Psi(\mathbf{e}_{s_i}, \mathbf{P}_{-i}) \cdot D_{i,s_i,s_i'} \\ &= \frac{1}{2} \sum_i \sum_{s_i < s_i'} p_{i,s_i}(t) p_{i,s_i'}(t) E_{i,s_i,s_i'} \cdot D_{i,s_i,s_i'} \geq 0\end{aligned} \quad (11)$$

where $D_{i,s_i,s_i'} = \psi_i(\mathbf{e}_{s_i}, \mathbf{P}_{-i}) - \psi_i(\mathbf{e}_{s_i'}, \mathbf{P}_{-i})$, $E_{i,s_i,s_i'} = \Psi(\mathbf{e}_{s_i}, \mathbf{P}_{-i}) - \Psi(\mathbf{e}_{s_i'}, \mathbf{P}_{-i})$, and the last inequality holds since from the condition of OPGs in (8), $D_{i,s_i,s_i'}$ and $E_{i,s_i,s_i'}$ always have the same sign. Since Φ is upper bounded and nondecreasing along the trajectories of the ODE, the convergence to an NE is guaranteed [8, Theorem 3.2, 3.3]. ■

While the convergence to an NE is guaranteed as $b \to 0$, a smaller value of b leads to a slower convergence rate. A proper value of b can be numerically determined to strike the desired tradeoff between the accuracy and rate of convergence for practical operations of the algorithm.

IV. Numerical Results

In order to evaluate the performance of the proposed scheme, we conduct a series of simulations. We consider a heterogeneous network in which there are 2 SPs each owning 3 channels. There are 10 SUs in the network, and the per-channel throughput is set to $R_{m,i} = \{2, 4, 6\}$ to reflect the modulation orders adopted under different RSS conditions. Fig. 2 shows the evolution of the choice probabilities of the actions (i.e., mixed strategy) for network selection using the proposed stochastic learning algorithm. With equal initial probabilities, it is observed that the network selection probabilities converge to pure strategies in around 300 and 100 cycles for $b = 0.2$ and $b = 0.5$, respectively. Note that SU #10 takes different strategies in the two cases. In Fig. 3, we test the deviation of the network selection of each of the 10 players. It is shown in Fig. 3(a) that when $b = 0.2$, unilateral deviation results in lower throughputs for all players, suggesting an NE point is reached by the learning algorithm. On the other hand, when $b = 0.5$, as shown in Fig. 3(b), SU #10 achieves a higher throughput by unilateral deviation, and thus the resulting strategy is not an NE point.

In Table II, we compare the performance of the proposed network selection scheme with two other approaches, namely, best RSS and (centralized) exhaustive search, which are described as follows:

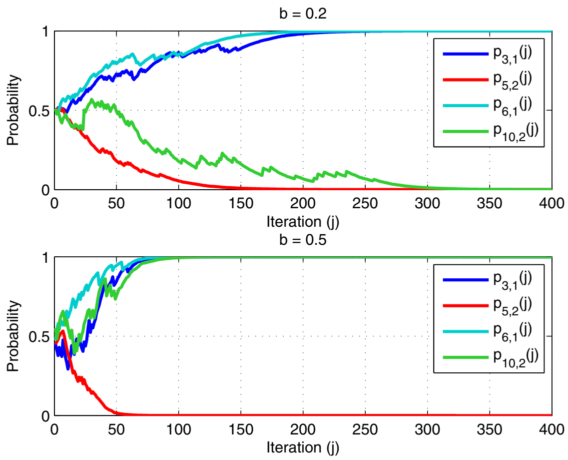

Fig. 2. Evolution of the mixed strategies (choice probability of actions) of some players, using different learning rates.

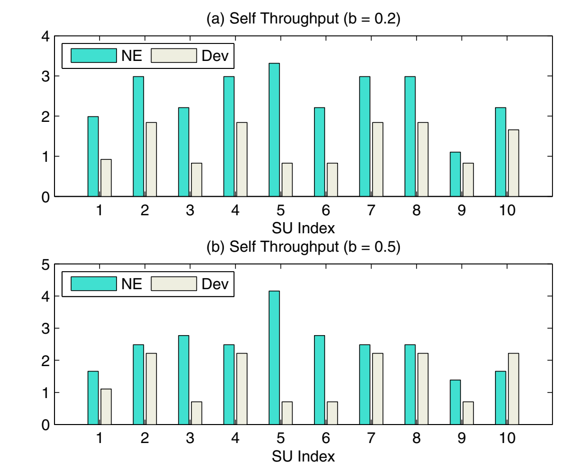

Fig. 3. Test of unilateral deviation from the resulting strategy profile of each of the 10 players, using different learning rates.

TABLE II
COMPARISON OF THE ACHIEVABLE EXPECTED SYSTEM THROUGHPUT OF THREE NETWORK SELECTION SCHEMES

	Proposed	Best RSS	Exhaustive
Scenario 1, u_{sum}	24.9662	24.4521	27.0621
Scenario 1, JFI	0.8974	0.7759	0.3822
Scenario 2, u_{sum}	25.9379	14.8554	25.9379
Scenario 2, JFI	0.9986	1.0000	0.8894

- In the best RSS scheme, each SU chooses the SP with the best per-channel throughput (i.e., $a_i = \arg\max_m R_{m,i}$). If there are more than one best SP, choose arbitrarily.
- In the exhaustive search, the channel availability statistics and the number of SUs are known to a centralized controller, and the action profile is selected so as to maximize the system throughput $u_{sum} = \sum_{i=1}^{N} u_i$.

The performance of different network selection schemes are evaluated by the system throughput u_{sum} and the fairness among SUs, measured by the Jain's fairness index (JFI), $J = u_{sum}^2/(N\sum_{i=1}^{N} u_i^2)$. We consider two scenarios for the simulation. In scenario 1, the SUs are randomly distributed. An SU may have better RSS from SP_1 ($R_{1,i} > R_{2,i}$), from SP_2 ($R_{1,i} < R_{2,i}$), or similar RSS from both SPs ($R_{1,i} = R_{2,i}$). In scenario 2, we set $R_{1,i} = 6$ and $R_{2,i} = \{2, 4\}, \forall i \in \mathcal{N}$. This describes a two-tier network where SP_1 is a small-cell serving indoor SUs, while SP_2 is a macro-cell located far apart. We observe that the efficiency of the learned NE strategy (ratio between u_{sum} of the proposed and exhaustive search methods) is above 90% for both scenarios. In addition, the exhaustive search method results in best u_{sum}, but suffers from poor fairness in scenario 1. This is due to the *winners-first* property of exhaustive search: If m can be found so that $R_{m,i} = 6$, SU_i is usually assigned to SP_m; on the other hand, those SUs with lower $R_{m,i}$ in both networks may be assigned to a less crowded SP instead of their own preference. The best RSS scheme has good system throughput in scenario 1 but not in scenario 2, since in this extreme case, all SUs are crowded in SP_1 and the resources of SP_2 are wasted. In contrast, the proposed method performs well in terms of both throughput and fairness under both scenarios. The results show the advantage of the proposed method: through the learning procedure towards equilibrium, the throughput of each SU is considered and the fairness can be maintained.

V. CONCLUSION

In this letter we have studied the problem of self-organized network selection in heterogeneous networks with time-varying channel availability and unknown number of secondary users. We formulated the network selection problem by an ordinal potential game. A decentralized stochastic learning-based algorithm has been proposed. Simulation results have demonstrated the convergence of the algorithm towards a pure strategy Nash equilibrium point. The proposed method outperforms the best RSS scheme in terms of average throughput, while the performance loss compared to the centralized exhaustive search is limited. Moreover, the proposed method achieves good fairness in various network scenarios.

REFERENCES

[1] M. Kassar, B. Kervella, and G. Pujolle, "An overview of vertical handover decision strategies in heterogeneous wireless networks," *Computer Commun.*, vol. 31, no. 10, pp. 2607–2620, Jun. 2008.
[2] K. Zhu, D. Niyato, P. Wang, and Z. Han, "Dynamic spectrum leasing and service selection in spectrum secondary market of cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1136–1145, Mar. 2012.
[3] M. Khan, H. Tembine, and A. Vasilakos, "Game dynamics and cost of learning in heterogeneous 4G networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 198–213, Jan. 2012.
[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] I. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr. 2008.
[6] D. Monderer and L. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.
[7] W. Zhong, Y. Xu, M. Tao, and Y. Cai, "Game theoretic multimode precoding strategy selection for MIMO multiple access channels," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 563–566, Jun. 2010.
[8] 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.