This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2010 proceedings.

Competitive Spectrum Access in Cognitive Radio Networks: Graphical Game and Learning

Husheng Li and Zhu Han

***Abstract*—Competitive spectrum access is studied for cognitive radio networks. Based on the assumption of rational secondary users, the spectrum access is modeled as a graphical game, in which the payoff of a secondary user is dependent on only other secondary users that can cause significant interference. The Nash equilibrium in the graphical game is computed by minimizing the sum of regrets. To alleviate the local knowledge of payoffs (each secondary user knows only its own payoff for different channels), a subgradient based iterative algorithm is applied by exchanging information across different secondary users. When information exchange is not available, learning for spectrum access is carried out by employing stochastic approximation (more specifically, the Kiefer-Wolfowitz algorithm). The convergence of both situations is demonstrated by numerical simulations.**

I. Introduction

In recent years, cognitive radio has emerged as a novel technique to alleviate the problem of inefficient frequency spectrum utilization [12]. In a cognitive radio system, a secondary user without license is allowed to access a licensed channel (there are typically multiple channels) if no primary user having license is using this channel; otherwise the secondary user should keep silent or look for another available channel since primary users have higher priority.

For a single-user cognitive radio system, the key problem is to find an available frequency channel, i.e. spectrum sensing. Once an idle channel is found, it is straightforward to complete the subsequent data transmission using this channel. However, for cognitive radio networks, it is possible that two or more secondary users sense the same channel and then collide if they all transmit over this channel (here we assume that orthogonal transmission is needed). Therefore, it is necessary to study how to allocate available resource to multiple secondary users in cognitive radio networks. One approach is to let potentially conflicting secondary users negotiate with each other and elect one to transmit over the idle channel. Theories like cooperative game, bargaining and welfare economics can be applied to achieve an efficient and fair resource allocation [4] [11] [16]. Another approach is to let secondary users learn the optimal strategies without information exchange [10].

Note that all these studies are based on the assumption that all secondary users can substantially interfere with each other and then cannot share the same channel. Using the terminology of game theory, the payoff of each player is dependent on the actions of all other players. This assumption may not be true in cognitive radio networks since two secondary users far away from each other may have little mutual impact. Therefore, *it is important to tackle the limited range of mutual impact in practical systems, which is not addressed in traditional game theoretic studies on cognitive radio systems.*

H. Li is with the Department of Electrical Engineering and Computer Science, the University of Tennessee, Knoxville, TN, 37996 (email: husheng@eecs.utk.edu). Z. Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX, 77004. This work was supported by the National Science Foundation under grants CyberTrust-0910461, CCF-0830451 and ECCS-0901425.

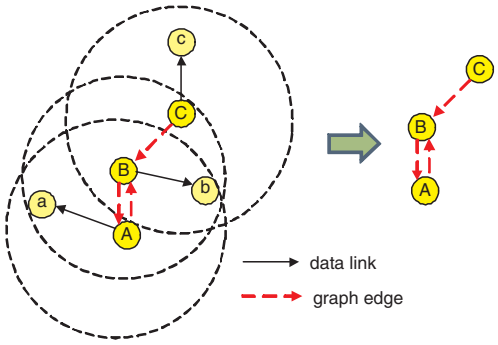

Fig. 1: Illustration of graphical game in cognitive radio networks.

In this paper, we apply a novel theory called *graphical games*, which was developed in recent years [1] [2] [5] [6] [7] [9] [15], to study the competition game of spectrum sensing and the corresponding learning in cognitive radio networks. In graphical games, the players are represented by nodes in a graph and the mutual impacts are denoted by edges. The payoff of each player is dependent on only its neighboring nodes (players). For the graphical game in cognitive radio networks, an example is shown in Fig. 1, where secondary transmitters/receivers are labeled by capital/small letters and the transmission range of each secondary transmitter is represented by a circle. We observe that competition exists between transmitters A and B while transmitter C has asymmetric impact on B.

Graphical games can well capture the limited range of mutual impact and is thus suitable for modeling a wide area cognitive radio network. To the authors' best knowledge, there have not been any studies on the application of graphical games in cognitive radio systems. We will adopt a random spectrum access scheme motivated by p-persistent carrier sense multiple access (CSMA), i.e. a secondary user transmits with probability p when an idle channel is sensed. Similar to the Aloha-like spectrum access scheme [10], there is no need for negotiation among secondary users, thus avoiding significant overhead. The possible actions of each secondary user are the selection of channels to sense. Then, the graphical

978-1-4244-6398-5/10/$26.00 ©2010 IEEE

game model will be used to compute the Nash equilibria using *no-regret learning*. When each secondary user knows its payoff of the game, but unknown to other secondary users, subgradient method will be used to minimize the sum of regrets in a distributed way. When the payoffs are unknown, learning algorithms are proposed for both cases with and without information exchange. For the case with information exchange, an incremental estimation of payoffs is applied, while *stochastic approximation* (the Kiefer-Wolfowitz algorithm) is used for the case without information exchange.

The remainder of this paper is organized as follows. System model will be introduced in Section II. The Nash equilibria will be computed in Section III and learning algorithms will be proposed in Section IV. Numerical results and conclusions are provided in Sections V and VI.

II. System Model

We consider a cognitive radio network consisting of N secondary users with an arbitrary network topology and M channels. The system is time slotted. Each time slot is divided into a spectrum sensing period, a communication period and an optional information exchange period. We assume that every secondary user can sense all channels but can transmit over only one channel. The selection of channel is based on the spectrum sensing results. An approach similar to p-persistent CSMA is used for the spectrum access, i.e. once a secondary user decides to use an available channel, it transmits with probability p and stays idle with probability $1-p$, where p is a constant. An adaptive p may improve the performance and can be learned from experience. However, the corresponding analysis is much more complicated and is beyond the scope of this paper.

For simplicity, we assume that all these N secondary users are affected by the same interruptions from primary users such that they share the same state of interruptions. This is reasonable since the interruption range of a primary user is typically much larger than the interference and transmission ranges of secondary users.

We denote by $\mathcal{I}_n$ the set of secondary users interfering secondary user n, i.e. if $i \in \mathcal{I}_n$, secondary users i and n cannot transmit simultaneously. We denote by $\mathcal{C}_n$ the set of secondary users that are able to communicate with secondary user n. Since the range of interference is typically larger than that of communication, we have $\mathcal{C}_n \subset \mathcal{I}_n$.

We do not explicitly model the interruption procedure of primary users. Since we assume that every secondary user can sense all channels, the game becomes one-snapshot and we only need to find strategies for the channel situation in each single time slot. When the secondary users can sense only a fraction of the channels, the system is partially observable and the spectrum sensing becomes a stochastic game. However, the stochastic game is complicated and is beyond the scope of this paper.

We assume that *orthogonality* is needed for data transmission, i.e. if two or more secondary users transmit over the same channel, the transmissions will collide and then fail. Therefore, the spectrum sensing is competitive. If secondary user n succeeds in transmitting over channel i, it receives a random payoff with expectation R_{ni}. Note that the expectation of payoff, R_{ni}, could be the probability of transmitting a packet successfully or the expectation of channel capacity. For generic purpose, we do not specify the exact meaning of R_{ni}.

The following assumptions are used throughout the whole paper.

- Every secondary user always has data to transmit, i.e. a full buffer model is used for the data traffic. It is interesting to study bursty data traffics. Then each secondary user also needs to learn the traffic patterns of other secondary users. However, it is beyond the scope of this paper.
- When the secondary users need to exchange information, we assume that a reliable common control channel is available. All corresponding information exchanges among neighboring users[1] can be completed in the information exchange period of each time slot.

III. Graphical Games of Spectrum Access

In this section, we first introduce the fundamental elements of the graphical game, i.e. graph, action, state and payoff. Then, we use no-regret approach to compute the equilibrium of the graphical game with communication constraint[2]. Throughout this section, we assume that every secondary user knows the expected payoffs $\{R_{nm}\}_{nm}$ of different channels for its transmission. This assumption will be relaxed in Section IV. However, every secondary user does not know the payoffs of all other secondary users.

A. Elements of Graphical Games

The elements of the graphical game are given below:

- Graph: we consider any arbitrary topology of the graph for competitive spectrum sensing.
- State: at each time slot, the system state is the set of idle channels found by spectrum sensing, denoted by S. Note that the actions of secondary users do not change the system state, which is determined by only the primary system.
- Action: the action of secondary user n, denoted by a_n, is the selection of channel. For user n and state S, we denote by u_{nm}^S the probability of choose channel m, which satisfies

$$\sum_{m \in S} u_{nm}^S = 1, \qquad \forall n = 1, ..., N. \tag{1}$$

We denote by vector $\mathbf{u}_n^S = \left(u_{nk_1}, ..., u_{nk_{|S|}}\right)$ the strategy of secondary user n, when $S = \left\{k_1, ..., k_{|S|}\right\}$, and by matrix $\mathbf{U}^S$ the strategy of all secondary users (called *global strategy*), where the n-th row is for the strategy vector of secondary user n.

[1] In this paper, two neighboring users mean users that can communicate with each other.

[2] Note that the no-regret algorithm is used to compute the correlated equilibrium of the game of spectrum access in [3], in which the computation is centralized.

- Payoff: when secondary user n chooses channel m, the expected payoff is given by

$$r_{nm}^S = pR_{nm} \prod_{l \in \mathcal{I}_n} \left(u_{lm}^S (1-p) + (1 - u_{lm}^S) \right), \quad (2)$$

since the decisions of different secondary users are independent.

Then, the expected payoff with respect to mixed strategy $\left\{u_{nm}^S\right\}$ of secondary user n is given by

$$\bar{r}_n^S = \sum_{m \in S} r_{nm}^S u_{nm}^S. \quad (3)$$

Without loss of generality, we assume that channels $1, ..., M$ are all idle in the time slot being considered, i.e. $S = \{1, ..., M\}$. For notational simplicity, we ignore the subscript of state S.

B. Regret Minimization

It is well known that the Nash equilibrium of a game means that, for any player, unilaterally changing its strategy incurs only payoff degradation. For the game of spectrum access, the Nash equilibrium $\{u_{nm}^*\}_{n,m}$ satisfies

$$\mathbf{u}_n^* = \arg\max_{\mathbf{u}_n} \sum_{j=1}^{M} u_{nj} r_{nj}, \quad (4)$$

where r_{nj} is the average reward when accessing channel j, which is given in (2). It is easy to verify that the solution of (4) is also a solution to the following optimization problem:

$$\arg\min_{\mathbf{U}} \sum_{n=1}^{N} \left(\max_{\mathbf{u}_n} \bar{r}_n(\mathbf{U}_{-n}, \mathbf{u}_n) - \bar{r}_n(\mathbf{U}) \right)^2, \quad (5)$$

where $\bar{r}_n(\mathbf{U})$ is the average payoff of secondary user n under strategy $\mathbf{U}$ and $\mathbf{U}_{-n}$ means the strategies of all secondary users except secondary user n. In many literatures, the term $\max_{\mathbf{u}} \bar{r}_n(\mathbf{U}_{-n}, \mathbf{u}_n) - \bar{r}_n(\mathbf{U})$ is called *regret*, which stands for the difference between the payoffs using the best response and the current strategy.

Then, to obtain a Nash equilibrium for the competitive spectrum access, secondary users can jointly optimize (5), which is the well known no-regret approach for computing Nash equilibrium. Note that we assume that *all secondary users follow the same protocol, e.g. they are designed and manufactured by the same company. Therefore, they have a common incentive to minimize the total regret.* We do not consider the possibility of attacks or selfishness, which is interesting but beyond the scope of this paper.

C. Distributed Regret Minimization

One approach to minimize the total regret in (5) is to let all secondary users to report their payoffs $\{R_{nm}\}_{n,m}$ to a processing center and then carry out a centralized optimization. In this paper, we apply a distributed approach using iterative computation, which is similar to the subgradient algorithm for distributed optimization [13] [14] [17]. Such an iterative approach is easier to implement and facilitates the distributed learning discussed in the next section.

We first initialize the strategies of all secondary users using arbitrary values. Then, in iteration t, neighboring secondary users exchange their current estimation of global strategies, denoted by $\mathbf{U}_n(t)$ for secondary user n[3]. Subsequently, each secondary user updates its own strategy to decrease its regret. For secondary user n, its optimal response to its estimation of global strategy $\mathbf{U}_n(t)$ is given by

$$\mathbf{u}_{nm}^* = \frac{1}{|\Omega_n(t)|}, \quad (6)$$

if

$$m = \arg\max_l r_{nl}(\mathbf{U}_n(t)), \quad (7)$$

where $\Omega_n(t)$ is the set of channels that maximizes the payoff in (7), and

$$\mathbf{u}_{nm}^* = 0, \qquad \text{otherwise.} \quad (8)$$

Obviously, the rule in (6) and (8) means that the optimal response is to choose the channel(s) that provides the optimal payoff subject to the strategies of all other secondary users. In practise, we also incorporate channel m such that $|r_{nm} - \max_l r_{nl}| \leq \delta$ into $\Omega_n(t)$ due to quantization errors in computation.

Then, secondary user n updates its own estimation of global strategy according to the optimal response and received global strategy estimations from neighboring users, which is given by

$$\begin{aligned} \mathbf{U}_n(t+1) &= \frac{1}{|\mathcal{C}_n| + 1} \sum_{l \in \mathcal{C}_n \bigcup \{n\}} \mathbf{U}_l(t) \\ &+ \alpha(t) \left(\mathbf{U}_l^*(t) - \mathbf{U}_l(t) \right), \end{aligned} \quad (9)$$

where $\mathbf{U}_l^*(t)$ is obtained by replacing the strategy of secondary user n in $\mathbf{U}_l(t)$ with $\mathbf{u}_n^*$ obtained in (6) and (8). $\alpha(t)$ is a step factor and we set $\alpha(t) = \frac{\alpha_0}{t}$. Note that the obtained strategy in (9) should be normalized to make the sum of probabilities unit.

The iteration continues until it converges and the convergence is guaranteed by the property of subgradient method [13] [14]. The procedure is summarized in Procedure 1.

Procedure 1 Iterative Computation of Nash Equilibrium

1: Initialize the strategies randomly.
2: **for** Each iteration **do**
3: Neighboring secondary users exchange their estimations of global strategy.
4: Each secondary user computes the optimal response to the current estimation of global strategy using (6) and (8).
5: Each secondary user updates its estimation of global strategy using (9) and normalizes the probabilities.
6: If some convergence condition is reached, stop.
7: **end for**

[3]Note that $\mathbf{U}_n$ is an estimation of the optimal strategies of all secondary users at secondary user n; therefore, it may be different for different secondary users.

IV. Learning for Spectrum Access

In Section III, we assume that every secondary user knows the expected payoffs of all channels with respect to itself, i.e. $\{R_{nm}\}_{m=1,...,M}$ for user n, but not the payoffs of other secondary users. However, in practical systems, this knowledge is unknown to secondary users at the beginning of network operation. To alleviate the challenge of unknown environment, we let secondary users learn the Nash equilibrium during the operation of network. Or equivalently, secondary users minimize the sum of regret in (5) incrementally. The major challenge is that the expression of (5) is unknown since the corresponding expected payoffs $\{R_{nm}\}_{n,m}$ are unknown. This means that we need to solve an optimization problem with unknown objective function. The corresponding solution is dependent on whether information exchange is allowed among secondary users. In this section, we discuss both cases with and without information exchange.

A. Learning with Information Exchange

When information can be exchanged among secondary users, we can still adopt the subgradient approach in Section III. The difficulty here is that the expected payoffs $\{R_{nm}\}_{n,m}$ are unknown to secondary users. This can be alleviated by estimating $\{R_{nm}\}_{n,m}$ during the learning procedure since $\{R_{nm}\}_{n,m}$ are assumed to be constants. The update of strategies is carried out every G rounds of spectrum accesses. Then, the learning procedure can be modified directly from Procedure 1 and is summarized in Procedure 2.

Procedure 2 Learning Procedure with Information Exchange Among Secondary Users

1: Initialize the strategy estimations randomly.
2: **for** Each update **do**
3: Carry out G rounds of the game.
4: Each secondary user estimates the expected payoffs according to the received payoffs during all previous rounds of the game.
5: Neighboring secondary users exchange their estimations of global strategy.
6: Each secondary user computes the estimation of optimal response to the current estimation of global strategy using (6) and (8), using the estimation of expected payoffs.
7: Each secondary user updates its estimation of global strategy using (9) and normalizes the probabilities.
8: If some convergence condition is reached, stop.
9: **end for**

B. Learning without Information Exchange

When there is no information exchange among secondary users, the regret minimization problem in (5) is much more challenging since the strategies of all secondary users change with time and the computation of optimal response cannot be done by simply estimating $\{R_{nm}\}_{n,m}$. In this paper, we adopt the Kiefer-Wolfowitz algorithm [8], which is used to optimize objective functions with unknown forms. Essentially, the Kiefer-Wolfowitz algorithm estimates the gradient along each direction from received payoffs and then updates the corresponding strategy along an ascending direction. The detail of the learning procedure is given below and is then summarized in Procedure 3.

1) Estimation of Gradient Vector: Before each update of strategies, we carry out multiple rounds of spectrum accesses for the secondary users to estimate the gradients with respect to different strategies. We let all secondary users change strategies in turn and estimate the corresponding gradient vector. Below is the procedure of estimating the gradient vector when it is secondary user n's turn to change its strategy. During this period, all other secondary users should keep their strategies fixed.

We denote by $\mathbf{e}_i$ the vector $(0, ..., 0, 1, 0, ...0)$, where only the i-th element is 1. For secondary user n, we carry out $2MG$ times of spectrum access (2G times for each channel). From the $2(m-1)G+1$-th to the $(2m-1)G$-th spectrum access, we change the strategy of secondary user n to

$$\mathbf{u}_n^{m+} = \frac{\mathbf{u}_n(t) + c_t \mathbf{e}_m}{\|\mathbf{u}_n(t) + c_t \mathbf{e}_m\|_1}, \tag{10}$$

where c_t is a step factor satisfying $c_t \to 0$ as $t \to \infty$. From the $(2m-1)G+1$-th to the $2mG$-th spectrum access, we change the strategy of secondary user n to

$$\mathbf{u}_n^{m-} = \frac{\mathbf{u}_n(t) - c_t \mathbf{e}_m}{\|\mathbf{u}_n(t) - c_t \mathbf{e}_m\|_1}. \tag{11}$$

Note that dividing by the 1-norms in (10) and (11) is to keep the sum of probabilities unit. We denote by R_n^{m+} and R_n^{m-} the average payoffs obtained for the two periods, respectively. Then, we define

$$y_{nm}(t) = \frac{R_n^{m+} - R_n^{m-}}{2c_t}. \tag{12}$$

After the $2MG$ spectrum accesses, we obtain an M-dimensional vector, which is given by

$$\mathbf{y}_n(t) = (y_{n1}(t), ..., y_{nM}(t)), \tag{13}$$

which is intuitively an estimation of the gradient vector of the regret of secondary user n.

2) Strategy Update: When all secondary users have estimated their gradient vectors, they update their strategies. For secondary user n, the strategy is updated as

$$\mathbf{u}_n(t+1) = \frac{\mathbf{u}_n(t) + \epsilon_t \mathbf{y}_n(t)}{\|\mathbf{u}_n(t) + \epsilon_t \mathbf{y}_n(t)\|_1}. \tag{14}$$

As we will see, the proposed learning procedure converges although we are still unable to provide a rigorous proof.

V. Numerical Results

In this section, we use numerical simulations to demonstrate the algorithms proposed in this paper. The network used for simulations si illustrated in Fig. 2, where solid lines mean that the two nodes are able to communicate with each other (thus can also interfere with each other) and dashed lines mean that the two nodes can interfere with each other but cannot communicate. For simplicity, we assume that the interference is symmetric although it may not be true in practical networks. We also assume that each real number in a message is quantized using q bits if secondary users exchange information.

Procedure 3 Procedure of Learning without Information Exchange

1: Initialize the strategies randomly.
2: **for** each update of strategies **do**
3: **for** each secondary user n **do**
4: fix the strategies of secondary users other than user n.
5: **for** each channel m **do**
6: Secondary user n computes $\mathbf{u}_n^{m+}$ and $\mathbf{u}_n^{m-}$ using (10) and (11), respectively.
7: Carry out $2G$ rounds of spectrum sensing, in which secondary user n uses $\mathbf{u}_n^{m+}$ and $\mathbf{u}_n^{m-}$ for rounds 1 to G and rounds $G+1$ to $2G$, respectively. Obtain R_n^{m+} and R_n^{m-}.
8: **end for**
9: Secondary user n computes the estimation of gradient vector, $\mathbf{y}_n(t)$, using (13).
10: **end for**
11: Secondary users update their strategies using (14).
12: **end for**

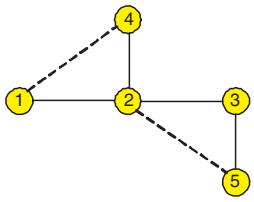

Fig. 2: Interference and communication graphs used for simulations.

A. Equilibrium Computing

We generate the payoffs using uniform random numbers between 0 and 1. For example, it could represent the probability of packet transmission success. We assume that the probability of transmission, p, equals 0.5.

In Figures 3 and 4, dynamics of the total regret and total payoff are shown with different numbers ofEach real number in a message is quantized using q bits. bits per probability value in messages exchanged among secondary users. We assume that $M = 8$, i.e. there are 8 available channels. We observe that the total regret converges monotonically close to 0, which means Nash equilibrium, after around 80 iterations while the total payoff is significantly improved. We also observe that the performance is insensitive to the number of bits used for quantizing each probability value since 4 bits are sufficient for one probability value.

B. Learning with Information Exchange

For the learning procedure with information exchange, the evolutions of total regret and total payoff are shown in Figures 5 and 6, respectively, versus different numbers of bits in each exchanged probability value. We set $G = 10$, i.e. 10 rounds of spectrum accesses are carried out before each update of strategies. Note that both curves are based on one realization; thus random fluctuations are unavoidable. We observe that, at the beginning of learning, there are substantial fluctuations due to the inadequacy of training data for estimating the expected payoffs $\{R_{nm}\}_{n,m}$. As more rounds of spectrum accesses are carried out, the estimations of $\{R_{nm}\}_{n,m}$ are more precise, thus making the total regret converge smoothly.

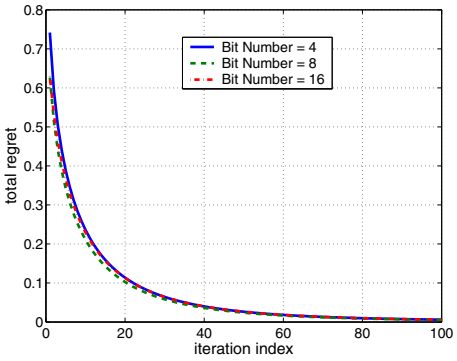

Fig. 3: Dynamics of total regret with respect to different numbers of bits in each probability value.

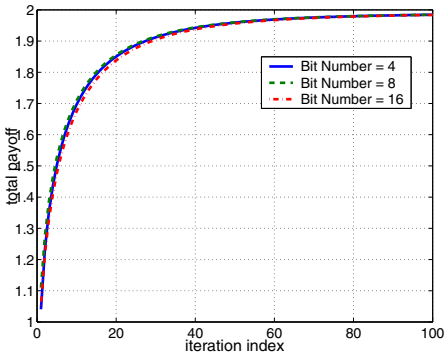

Fig. 4: Dynamics of total payoff with respect to different numbers of bits in each probability value.

C. Learning without Information Exchange

The dynamics of total regret and total payoff are given in Figures 7 and 8, respectively, for the procedure of learning without information exchange. We set $G = 1000$, i.e. 1000 rounds of spectrum accesses are carried out for each secondary user before updating its strategy. Simulation shows that, when

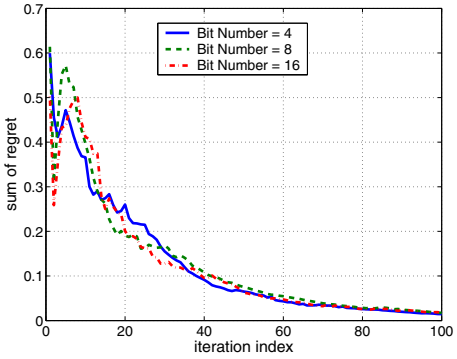

Fig. 5: Dynamics of total regret with respect to different numbers of bits for learning with information exchange.

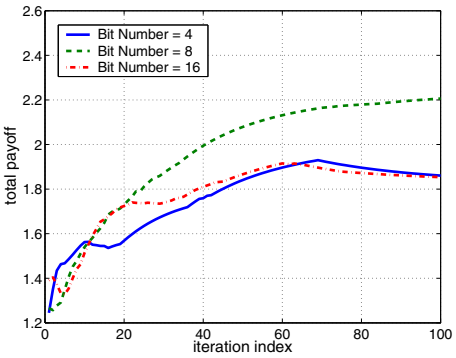

Fig. 6: Dynamics of total payoff with respect to different numbers of bits for learning with information exchange.

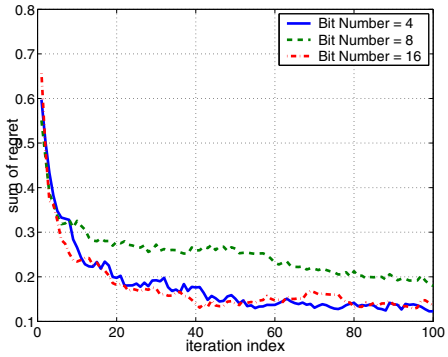

Fig. 7: Dynamics of total regret with respect to different numbers of bits for learning without information exchange.

using smaller G, the learning procedure does not converge. We observe that the learning procedure converges despite some fluctuations. However, the learning procedure does not converge to zero-regret and the convergence is very slow: it takes around 6×10^5 rounds of spectrum accesses to converge. If each spectrum access costs 1ms, it takes 10 minutes to converge. Therefore, further studies are necessary to accelerate the learning procedure, e.g. letting secondary users in an independent set, i.e. secondary users in this set do not interfere with each other, to learn simultaneously.

VI. Conclusions

We have studied the competitive spectrum access in cognitive radio networks. In contrast to conventional game theoretical studies on cognitive radio systems, we have modeled the spectrum access as a graphical game with arbitrary topology, in which a secondary user is affected by only nearby secondary users, instead of the whole network. To compute the Nash equilibrium in the game, we have applied the subgradient algorithm for minimizing the total regret in a distributed manner. When the payoffs are unknown to secondary users and there is no information exchange, we have incorporated stochastic approximation (Kiefer-Wolfowitz algorithm) into the procedure of learning for spectrum access. Numerical

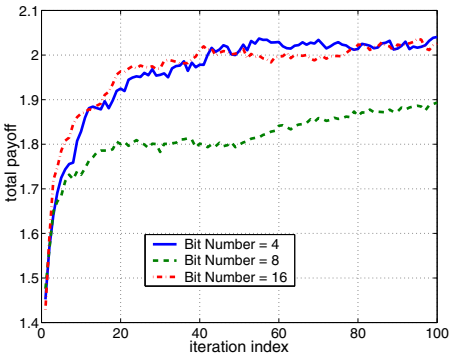

Fig. 8: Dynamics of total payoff with respect to different numbers of bits for learning without information exchange.

simulation results have demonstrated the convergence and asymptotic performance of the proposed algorithms.

References

[1] C. Daskalakis and C. Papadimitriou, "Computing pure Nash equilibria in graphical games via Markov random fields," in *Proc. of the 7th ACM Conferene on Electrionic Commerce*, 2006.

[2] E. Elkind, L. Goldberg and P. Goldberg, "Graphical games on tree revisited," in *Proc. of the 7th ACM Conferene on Electrionic Commerce*, 2006.

[3] Z. Han, C. Pandana and K. J. R. Liu, "Distributive opportunistic spectrum access for cognitive radio using correlated equilibrium and no-regret learning," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, 2007.

[4] E. Hossain, D. Niyato and Z. Han, *Dynamic Spectrum Access in Cognitive Radio Networks*. Cambridge University Press, UK, 2009.

[5] S. Kakade, M. Kearns, J. Langford and L. Ortiz, "Correlated equilibria in graphical games," in *Proc. of the 4th ACM Conference on Electronic Commerce (EC)*, 2003.

[6] S. Kakade, M. Kearns and L. Ortiz, "Graphical economics," in *Proc. of Conference on Learning Theory (COLT)*, 2004.

[7] S. Kakade, M. Kearns and L. Ortiz, R. Pemantle and S. Suri, "Economic properties of social networks," in *Proc. of the 18th Annual Conference on Neural Information Processing Systems (NIPS)*, 2004.

[8] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Statist.*, vol.23, pp.462–466, 1952.

[9] M. L. Littman, M. Kearns and S. Singh, "An efficient exact algorithm for singly connected graphical games," in *Proc. of the 15th Annual Conference on Neural Information Processing Systems (NIPS)*, 2001.

[10] H. Li, "Multi-agent Q-Learning of channel selection in multi-user cognitive radio systems: A two by two case," in *Proc. of IEEE Internatinal Conference on Systems, Man and Cybernetics (SMC)*, 2009.

[11] H. Liu, A. B. MacKenzie and B. Krishnamachri, "Bargaining to improve channel sharing between selfish cognitive radios," preprint.

[12] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Proc. IEEE Int. Workshop Mobile Multimedia Communications*, pp. 3–10, 1999.

[13] A. Nedić and D. P. Bertsekas, "Convergence rate of incremental subgradient algorithms," *Stochastic Optimization: Algorithms and Applications*, Kluwer Academic Publishers, 2000.

[14] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," LIDS Technical Report 2755, MIT, Lab, 2007.

[15] L. Ortiz and M. Kearns, "Nash propagation for loopy graphical games," in *Advances in Neural Information Processing Systems*, MIT Press, 2003.

[16] B. Wang, Y. Wu, Z. Ji, K. J. R. Liu and T. C. Clancy, "Game theoretical mechanism design for cognitive radio network with selfish users," *IEEE Signal Processing Magazine*, Vol.25, pp.74–84, Nov. 2008.

[17] J. N. Tsitsiklis and Z. Luo, "Communication complexity of convex optimization," *Journal of Complexity*, Vol.3, pp.231–243, 1987.