Self-Learning Repeated Game Framework for Distributed Primary-Prioritized Dynamic Spectrum Access

Beibei Wang, Zhu Ji*, and K. J. Ray Liu

Department of Electrical and Computer Engineering and Institute for Systems Research,
University of Maryland, College Park, MD 20742, USA
* Qualcomm, San Diego, CA 92121, USA

Abstract— Dynamic spectrum access has become a promising approach to fully utilize the scarce spectrum resources. In a dynamically changing spectrum environment, it is very important to design a distributed access scheme that can coordinate different users' access adapt to spectrum dynamics with only local information. In this paper, we propose a self-learning repeated game framework for distributed primary-prioritized dynamic spectrum access through modeling the interactions between secondary users as a noncooperative game. With the proposed framework, the inefficiency due to users' selfish behavior can be highly improved, and the secondary users can distributively obtain their optimal access probabilities with only local observations. The simulation results show that the proposed framework can achieve comparable performances with those of the centralized primary-prioritized dynamic spectrum access scheme.

I. INTRODUCTION

With the dramatic development of the mobile telecommunication industry in the last decades, the demand for wireless spectrum resources has been growing rapidly. However, the inflexible spectrum policies by FCC result in a large portion of the scarce spectrum resources remaining unutilized. The inefficient usage of the limited spectrum necessitates recent development of dynamic spectrum access techniques [1], [2], through the use of *cognitive radio* technology [3]. By exploiting the spectrum in an opportunistic fashion, dynamic spectrum access enables the secondary users to sense which portions of the spectrum are available, select the best channel, coordinate access to spectrum channels with other users and vacate the channel when a licensed user appears.

In order to fully utilize the limited spectrum resources, efficiently and fairly sharing the spectrum among multiple dissimilar secondary users becomes an important issue. There are several previous efforts addressing this issue, on an opportunistic basis [4]-[6], or a negotiated basis [7]-[12]. In [4], a novel random access protocol by using agile radios was proposed to "pack" all the radio systems tightly together in the spectral domain, and achieve airtime fairness between dissimilar secondary users. The work in [5] examined the impact of secondary user access patterns on achievable improvement in spectrum utilization, and proposed a feasible spectrum sharing scheme. The authors in [6] proposed a primary-prioritized Markov dynamic spectrum access to optimally coordinate secondary users' spectrum access and achieve a good tradeoff between efficiency and fairness statistically.

In order to study the users' selfish behaviors for distributed dynamic spectrum access in wireless networks, game theoretical approaches have been developed for dynamic spectrum access. A local bargaining mechanism was proposed in [7] to distributively optimize the efficiency of spectrum allocation and maintain bargaining fairness among secondary users. In [8], auction mechanisms were proposed for sharing spectrum among multiple users such that the interference was below certain level. In [9], the authors proposed a repeated game approach to enlarge the set of achievable rates, in which the spectrum sharing strategy could be enforced by the Nash Equilibrium of dynamic games. In [10]-[12], belief-based dynamic pricing approaches were developed to optimize the overall spectrum efficiency while keeping the participating incentives of the selfish users based on double auction rules.

Although existing dynamic spectrum access schemes have successfully enhanced spectrum efficiency, in order to achieve more flexible spectrum access in long-run scenarios, some basic questions still remain unanswered. First, considering the constant spectrum environment variations and lack of centralized authorities, the spectrum access scheme should be able to distributively adapt to the spectrum dynamics due to channel variations, varying traffic load, and user mobility with only local observations. From the game theoretical point of view, the spectrum access scheme needs to be designed as a multi-stage repeated game framework instead of a static

game. Furthermore, the users may belong to different authorities and aim to maximize their own interests, so they have no incentives to cooperate with each other, and the noncooperative behavior may result in inefficient and unfair outcomes. Therefore, novel spectrum access schemes should be developed to enhance the efficiency and fairness of the game.

Motivated by the preceding, in this paper we propose a self-learning repeated game framework for distributed primary-prioritized dynamic spectrum access [6]. Considering multiple secondary users coexist and constantly access some temporarily unused licensed spectrum band, we model the spectrum access as a noncooperative game. To solve the inefficiency of static game Nash Equilibrium (NE) due to secondary users' selfish behavior, we develop an efficient self-learning repeated game approach to enhance the game to a cooperating point with only local observations. The contributions of the proposed repeated game framework are multi-fold. First, by modeling the spectrum access process as a self-learning repeated game, we are able to optimally coordinate secondary users' access adapt to the spectrum dynamics with only local information. More importantly, the spectrum efficiency is improved compared to the static game NE. Moreover, the proposed distributed repeated game approach can achieve comparable performances with those of the centralized proportional-fair dynamic spectrum access [6], therefore, provide better fairness among dissimilar secondary users than the static game NE along with higher spectrum efficiency.

The remainder of this paper is organized as follows: The system model of the dynamic spectrum access is described in Section II. The self-learning repeated game approach for the dynamic spectrum access is proposed in Section III. The simulation studies are provided in Section IV. Finally, Section V concludes this paper.

II. SYSTEM MODEL

We consider dynamic spectrum access networks where multiple secondary users are allowed to access the temporarily unused licensed spectrum bands on an opportunistic basis, without conflicting or interfering the primary spectrum holders' usage. Such scenarios can be envisioned in many applications. Considering the fact that heavy spectrum utilization often takes place in unlicensed bands while licensed bands often experience low (e.g., TV bands) or medium (e.g., some cellular bands) utilization, IEEE 802.22 [13] proposes to reuse the fallow TV spectrum without causing any harmful interference to incumbents (i.e., the TV receivers). With
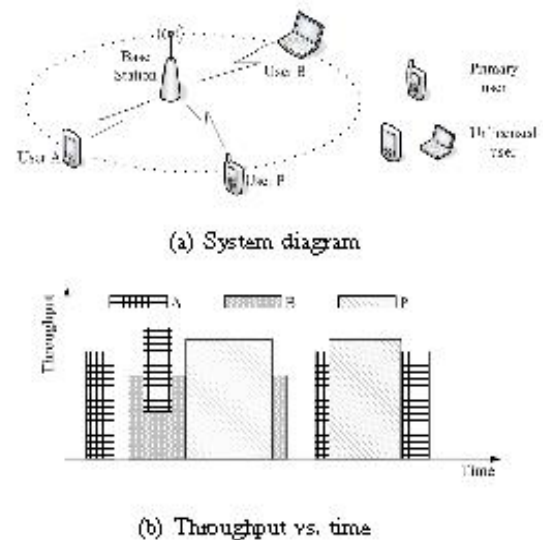

(a) System diagram

(b) Throughput vs. time

Fig. 1: System model.

the development of cognitive radios, ancillary services carried by the Ultra-High Frequency (UHF) television transmission are technically feasible. Moreover, with regard to more efficient utilization of some cellular bands, in [14], it is proposed to share the spectrum between a cellular communication system and wireless local area network (WLAN) systems. In rural areas where there is little demand on the cellular communication system, the WLAN users can efficiently increase their data rates by sharing the spectrum.

In order to take advantage of the temporally unused spectrum holes in the licensed band, without loss of generality we consider a snapshot of the above spectrum access networks shown in Fig. 1, where two secondary users and one primary user coexist, and the secondary users opportunistically utilize spectrum holes in the licensed band. Note that the system diagram shown here serves only as an example model and can be extended to the scenario with multiple secondary users in a similar way.

Let's denote the primary user by P, and the secondary users by A and B. For each user γ, where $\gamma \in \{A, B, P\}$, its offered traffic is modeled with two independent Poisson processes, with the service-request rate λ_γ and the departure rate μ_γ.

Since the primary user's spectrum usage in its licensed band should not be affected by the operation of any other secondary user [15] [16], we assume that once primary user P appears, any secondary user should stop transmission, buffer their interrupted traffic, continue scanning the licensed band, and immediately resume transmission once the licensed band becomes idle again.
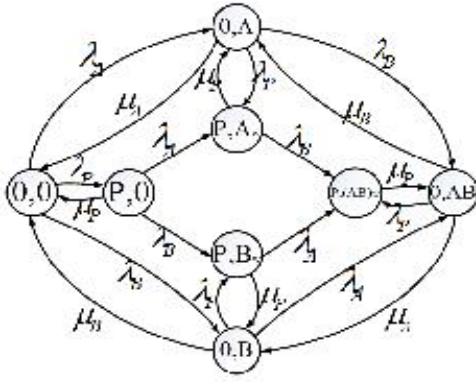
Fig. 2: The rate diagram of PP-CTMC.

TABLE I: The Eight States of PP-CTMC

Index	State	Description
0	$(0,0)$	Spectrum is idle
1	$(0,B)$	Secondary user B is in service
2	$(0,A)$	Secondary user A is in service
3	$(0,AB)$	Both A and B are in service
4	$(P,0)$	Primary user P is in service
5	(P,B_w)	P is in service; B is waiting
6	(P,A_w)	P is in service; A is waiting
7	$(P,(AB)_w)$	P is in service; A and B are waiting

Moreover, if more than one secondary users are allowed to share the licensed frequency-band, the efficiency of spectrum usage can be further improved. However, too much coexistence of the secondary users may result in severe mutual interference, and we will describe later how to coordinate the access of secondary users to alleviate the interference level.

From the preceding assumptions, it is clear that the spectrum environment is highly dynamic. The conventional spectrum allocation schemes specific for a static spectrum environment are no longer feasible. This is because using global optimization approaches, after each change in the spectrum environment, e.g., a secondary user enters or leaves some frequency band, the network needs to re-optimize the spectrum allocation for all users completely, resulting in high computational complexity. Therefore, in order to characterize the dynamics of the spectrum environment and fully utilize the statistics of different users' spectrum access, we model the interactions between the primary user and the secondary users as a Primary-Prioritized Continuous-Time Markov Chain (PP-CTMC), illustrated in Figure 2. The states of PP-CTMC are described in Table I.

Assume at first the licensed spectrum band is idle,

i.e., PP-CTMC is in state $(0,0)$. The two secondary users contend to operate in the spectrum. Upon the first access attempt of some user, say user A, PP-CTMC enters state $(0,A)$ with transition rate λ_A. If user A finishes its service duration before any other user requests spectrum access, PP-CTMC then transits to state $(0,0)$ with rate μ_A. If user B's service request arrives before A completes its service, PP-CTMC transits to state $(0,AB)$ with rate λ_B, where both secondary users share the spectrum usage using Code-Division Multiple Access (CDMA) techniques. Once user B (or A)'s service duration is completed, PP-CTMC transits from state $(0,AB)$ to $(0,A)$ (or $(0,B)$), with rate μ_B (or μ_A).

However, primary user P may, once in a while, appear during the service duration of the secondary users, i.e., when PP-CTMC is in state $(0,A)$, $(0,B)$ or $(0,AB)$. Suppose the licensed band is being occupied by user A. If user A detects that primary user P needs to acquire the spectrum band, A ceases its transmission, buffers its interrupted traffic, and keeps sensing the band until P finishes operating in the band. Therefore, PP-CTMC transits from state $(0,A)$ to (P,A_w) with rate λ_P. If primary user P finishes its service before B's access, A will continue its transmission, and PP-CTMC transits from state (P,A_w) to $(0,A)$ with rate μ_P. In contrast, if the access request of B arrives before primary user P completes its service duration, B also buffers its traffic, and PP-CTMC transits to state $(P,(AB)_w)$ with rate λ_B. In state $(P,(AB)_w)$, both A and B keep sensing the spectrum. Once P is sensed to vacate, PP-CTMC transits to state $(0,AB)$ with rate μ_P, where A and B share the spectrum band. Also, when PP-CTMC is in state $(P,0)$, if secondary users attempt to access the spectrum, they are kept sensing until P finishes its service, and PP-CTMC transits to state (P,A_w) or (P,B_w), with rate λ_A or λ_B, respectively.

The equation array governing the above system is given by

$$(\lambda_A + \lambda_B + \lambda_P)\Pi_0 = \mu_B\Pi_1 + \mu_A\Pi_2 + \mu_P\Pi_4, \quad (1)$$

$$(\mu_B + \lambda_P + \lambda_A)\Pi_1 = \lambda_B\Pi_0 + \mu_A\Pi_3 + \mu_P\Pi_5, \quad (2)$$

$$(\mu_A + \lambda_P + \lambda_B)\Pi_2 = \lambda_A\Pi_0 + \mu_B\Pi_3 + \mu_P\Pi_6, \quad (3)$$

$$(\mu_A + \mu_B + \lambda_P)\Pi_3 = \lambda_A\Pi_1 + \lambda_B\Pi_2 + \mu_P\Pi_7, \quad (4)$$

$$(\mu_P + \lambda_A + \lambda_B)\Pi_4 = \lambda_P\Pi_0, \quad (5)$$

$$(\mu_P + \lambda_A)\Pi_5 = \lambda_P\Pi_1 + \lambda_B\Pi_4, \quad (6)$$

$$(\mu_P + \lambda_B)\Pi_6 = \lambda_P\Pi_2 + \lambda_A\Pi_4, \quad (7)$$

$$\mu_P\Pi_7 = \lambda_B\Pi_6 + \lambda_P\Pi_3 + \lambda_A\Pi_5, \quad (8)$$

$$\Pi_0 + \Pi_1 + \Pi_2 + \Pi_3 + \Pi_4 + \Pi_5 + \Pi_6 + \Pi_7 = 1. \quad (9)$$

The stationary probability of each state is solvable from the above linear equation array, since the corresponding coefficient matrix can be verified nonsingular.

As the traffic of the secondary users becomes heavier, they have more chances to coexist in the band, and suffer more throughput degradation due to mutual interference. To coordinate secondary users' access and manage the interference, a centralized control is often needed. This approach, however, requires global information from all users. Thus, in the following we will develop a distributed game theoretical framework to optimize the system performance with only local information.

III. SELF-LEARNING REPEATED GAME FRAMEWORK

According to the system model, in this section, we first describe a centralized optimal dynamic spectrum access problem. Then, we formulate the spectrum access problem as a noncooperative game to study secondary users' selfish behaviors, and further develop a distributed repeated game framework to enhance the efficiency of the game equilibrium.

A. Centralized Optimal Dynamic Spectrum Access

One of the most important goals in spectrum sharing is efficient spectrum utilization, i.e., high throughput achieved by each secondary user through successful acquisition of a spectrum band. From a statistical point of view, the secondary users want to maximize their average throughput. Given the system equation array in Section II, we can obtain the stationary state probabilities, denoted by Π_{s_i}, where $s_i \in \{(0, A), (0, B), (0, AB)\}$. Since Π_{s_i} can be equivalently viewed as the ratio of allocation time to state s_i to the entire reference time, the product of Π_{s_i} and the capacity that secondary user γ achieves when operating in state s_i represents one average throughput component acquired by user γ in state s_i. Therefore, we can express the total average throughput for user γ as follows,

$$U_\gamma = \Pi_{(0,\gamma)} r_1^\gamma + \Pi_{(0,AB)} r_2^\gamma, \qquad (10)$$

where r_1^γ and r_2^γ are channel capacities for user γ when it operates in the licensed band alone and with the other secondary user, respectively.

In order to alleviate the mutual interference among secondary users, we introduce the spectrum access probability for user A and user B, denoted by a_A and a_B, respectively. Then, the resulting random access process can be approximated by slightly modifying the original CTMC. Because each secondary user γ's traffic is admitted with probability a_γ, the actual arrival rate is

approximated by $a_\gamma \lambda_\gamma$. Therefore, the system optimization goal is to determine a_A and a_B, such that the utility function of the system can be maximized, i.e.,

$$\{a_\gamma\} = \arg \max_{0 \le a_\gamma \le 1} U_{sys}(\{a_\gamma\}), \qquad (11)$$

where $\forall \gamma \in S \triangleq \{A, B\}$.

Since a good spectrum sharing scheme not only can efficiently utilize the spectrum resources, but also can provide fairness among different users, the centralized optimal dynamic spectrum access scheme can be defined to maximize the average throughput based on the proportional-fairness (PF) criterion. Thus, in (11), $U_{sys}(a_A, a_B)$ can be written as [6]

$$U_{sys}(a_A, a_B) = \prod_{\gamma \in S} U_\gamma(a_A, a_B). \qquad (12)$$

In order to solve (11), a centralized management point is required to collect all the information about the traffic statistics and the channel capacities from the secondary users, which is difficult to implement in a real spectrum access scenario. Therefore, in the next we will develop game theoretical approaches to solve this problem with only local information.

B. Spectrum Access Game and Nash Equilibrium

Different secondary users may belong to different authorities, so it is natural to assume that each secondary user γ selfishly selects its access probability a_γ to optimize its own throughput. In this case, game theory [17] is a successful approach to study the interactions between selfish secondary users.

Let $G = \{S, (A_\gamma)_{\gamma \in S}, (U_\gamma)_{\gamma \in S}\}$ be a finite spectrum access game in strategic form, where S is the set of players, A_γ is the strategy space for player γ, and U_γ is the utility function for player γ. Then, in our spectrum access game, we have $S = \{A, B\}$, $A_\gamma = [0, 1]$, and U_γ is defined in (10) with all λ_γ's replaced by $a_\gamma \lambda_\gamma$'s. Let's denote the action for user γ and its opponent as a_γ and $a_{-\gamma}$, respectively, and define $A_{-\gamma}$ as the strategy space for user γ's opponents. Then, the Nash Equilibrium (NE), which states that in the equilibrium each user selects the best response strategy given that the other users are playing the Nash Equilibrium as well, is defined as follows,

Definition 1: A strategy profile $a^* = (a_\gamma^*, a_{-\gamma}^*)$ is a Nash Equilibrium if, $\forall \gamma \in S$, $\forall a_\gamma$, $a_\gamma^* \in A_\gamma$, $a_{-\gamma}^* \in A_{-\gamma}$,

$$U_\gamma(a_\gamma^*, a_{-\gamma}^*) \ge U_\gamma(a_\gamma, a_{-\gamma}^*). \qquad (13)$$

Since the spectrum occupation time ratios $\Pi_{(0,\gamma)}$ and $\Pi_{(0,AB)}$ are increasing with the access probability a_γ,

TABLE II: Utility table of the spectrum access game

User A \ User B	Cooperate(PF)	Non-Cooperate
Cooperate(PF)	$(U_A^{(c,c)}, U_B^{(c,c)})$	$(U_A^{(c,n)}, U_B^{(c,n)})$
Non-cooperate	$(U_A^{(n,c)}, U_B^{(n,c)})$	$(U_A^{(n,n)}, U_B^{(n,n)})$

TABLE III: Self-Learning Repeated Game Algorithm

Initialization: $t = 0$
$\quad a_\gamma^t = a_{min}, \forall \gamma$. Choose small increment $\vec{\xi}, \eta$.
Iteration: $t = 1, 2, \cdots$
$\quad$ Observe $U_\gamma^{t-1}(\vec{a}_\gamma^{t-1})$ and $U_\gamma^{t-1}(\vec{a}_\gamma^{t-1} + \vec{\xi})$,
$\quad$ Obtain
$\quad \Delta U_\gamma^{t-1} = U_\gamma^{t-1}(\vec{a}_\gamma^{t-1} + \vec{\xi}) - U_\gamma^{t-1}(\vec{a}_\gamma^{t-1})$,
$\quad$ For each γ such that $\Delta U_\gamma^{t-1} > 0$,
$\quad\quad a_\gamma^t = a_\gamma^{t-1} + \eta \frac{\Delta U_\gamma^{t-1}}{U_\gamma^{t-1}(a_\gamma^{t-1})}$,
$\quad\quad a_\gamma^t = \min(a_\gamma^t, 1)$.
End when: No improvement in U_γ.

we can verify that $\partial U_\gamma / \partial a_\gamma > 0$. Thus, the NE for the spectrum access game is $a_\gamma = 1, \forall \gamma$. By greedily accessing the licensed band, each secondary user increases its occupation time duration, so as to boost its average throughput. However, if the secondary users are located very close to each other, such a greedy access pattern will greatly increase the probability that both the secondary users coexist in the spectrum band, and thus heavily degrade each user's throughput due to severe mutual interference. When more than two secondary users compete for the licensed spectrum, the throughput each secondary user can obtain will become even worse if all of them access the spectrum with $a_\gamma = 1$. Consider an example illustrated in Table II, where the "Cooperate(PF)" represents the secondary users access the spectrum band with the optimal probability a_γ^{opt} through solving (11) with PF criterion, the "Non-Cooperate" represents the access probabilities for both secondary users are 1, with $U_A^{(c,n)} < U_A^{(n,n)} < U_A^{(c,c)} < U_A^{(n,c)}$, and $U_B^{(n,c)} < U_B^{(n,n)} < U_B^{(c,c)} < U_B^{(c,n)}$. We can observe that when the secondary users cooperate with each other and access the spectrum band less aggressively, both of them can get higher utilities than those in the NE. Therefore, it is not efficient for the secondary users to play NE in the spectrum access game. To enhance the game to an efficient cooperating point rather than the NE, we propose a self-learning repeated game framework in the next.

C. Self-Learning Repeated Game Framework

From the previous discussion, the NE of the static spectrum access game is $a_\gamma = 1, \forall \gamma$, which is inefficient. In practice, however, the secondary users have continuous spectrum access requests as long as the primary user does not occupy the licensed band. Therefore, the game is played more than once and can be modeled as a *repeated game*. The definition of a *repeated game* is stated as follows [18].

Definition 2: If we denote G as a static spectrum access game and δ as a discount factor, the T-period *repeated game*, denoted by $G(T, \delta)$, consists of game G repeated T times. The average payoff to user γ is the expected sum of stage game utilities discounted at rate

δ, represented by

$$\bar{U}_\gamma^T = (1 - \delta) \sum_{t=1}^{T} \delta^{(t-1)} U_\gamma(\vec{a}(t)), \qquad (14)$$

where $U_\gamma(\vec{a}(t))$ represents the utility of user γ at each stage game played at time t.

Since the secondary users always have continuous spectrum access requests with high probability, without loss of generality, we can further model the game as a repeated game with infinite horizon. Unlike a game played once, a repeated game allows for a strategy to be contingent on past moves, thus allowing for reputation effects and retribution. The player's payoff in a repeated game is a discounted summation of her/his payoff at each stage. One of the most important results in repeated game theory is **Folk Theorem** [18], which asserts that for infinite repeated games there exists a discount factor $\hat{\delta} < 1$ such that any feasible and individually rational payoff can be enforced by an equilibrium for any discount factor $\delta \in (\hat{\delta}, 1)$. Thus, by playing a strategic game many times, more efficient Nash Equilibria can be achieved in a repeated game framework.

Further, if we assume in the infinitely repeated game, the secondary users are aware that whenever a noncooperation is detected, all the other cooperative secondary users will turn to be noncooperative, thus resulting in a global performance decrease, it is better for the secondary users to conform with the cooperating point. Therefore, we can develop a self-learning algorithm for the infinitely repeated game, by which the secondary users can learn the optimal access probabilities in a distributed way. The heuristics of the self-learning algorithm is stated as follows. With only local information, the secondary users can only observe the history of their own actions and the corresponding utility functions.
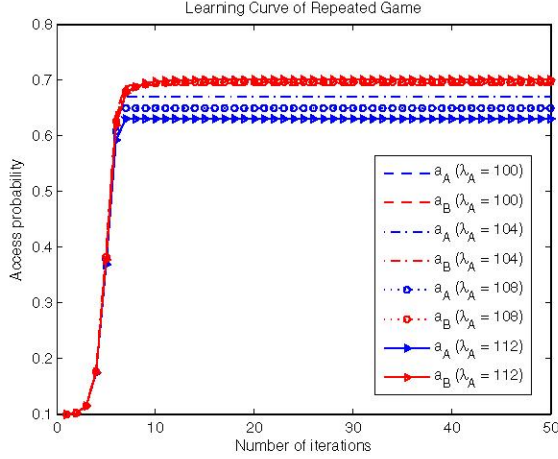
Fig. 3: Learning curves for the repeated game.

Fig. 4: Comparison of the learned access probabilities and the theoretical optima vs. λ_A.

Therefore, the best way for the secondary user to learn the optimal access probability is to observe how it will affect its own utility value to change the access probability. Assume under the threat of potential punishment, the secondary users simultaneously increase the access probabilities from very small values (e.g., 0.1) gradually. If the new utility is greater than the previous one, the new access probability will be adopted. Otherwise, the previous access probability is kept. The proposed algorithm is shown in Table III. We note that the access probability is increased proportionally to the increase in the utility function ΔU_γ^{t-1}; however, in order to keep the updated access probability bounded, normalization factor needs to be employed here. Therefore, we introduce $U_\gamma^{t-1}(a_\gamma^{t-1})$, the utility obtained before changing the access probability, as the denominator of ΔU_γ^{t-1}. The increase in the access probability depends on the "step size" η and the normalized increment in the secondary user's utility function. The access probabilities are increased iteratively until no improvement in the utilities can be achieved. Note that the algorithm showed in Table III can also be employed to the scenario with more than two secondary users by carefully choosing the increment $\vec{\xi}$ and η. And we will show by simulations in the next section that the self-learning process will converge to a near optimal cooperating point within several iterations.

IV. SIMULATION RESULTS AND ANALYSIS

In order to evaluate the performance of the proposed self-learning repeated game approach for the primary-prioritized dynamic spectrum access scheme, we first study the convergence of the learning algorithm. Then, we co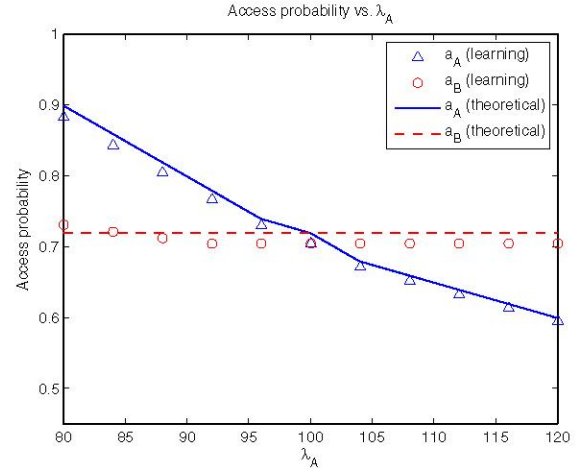mpare the results of the self-learning algorithm with the theoretical optimal solutions through solving (11) with the PF criterion.

In the simulations, the communication bandwidth is 200 KHz, the transmission power is $p_\gamma = 2$ mW, and the thermal noise power is $n_0 = 10^{-15}$ W. The propagation loss factor is 3.6. The transmitter of user A is at $(0m, 0m)$, and its receiver is at $(200m, 0m)$. The transmitter of user B is at $(200m, 120m)$, and its receiver is at $(0m, 120m)$. The service rates of P, A and B are all set to be 100 ms^{-1}. The spectrum access request rates of P and B are chosen as 85 ms^{-1} and 100 ms^{-1}, while the spectrum access request rate of A varies from 80 to 120 ms^{-1}. In the self-learning algorithm, we choose $\vec{\xi} = [5 \times 10^{-4}; 5 \times 10^{-4}]$, and $\eta = 0.8$.

In Figure 3, we show the learning curves of both user A and B when λ_A equals to different values, and in Figure 4 we show the comparison of the learned access probabilities with the theoretical optimal probabilities. We can see that the self-learning repeated game algorithm converges to near optimal solutions with about 10 iterations. The fast convergence of the learning algorithm ensures the secondary users to obtain proportional-fair utility distributions before the licensed band is required by the primary user. From Figure 4, we can see that the learning results of the repeated game is very close to the PF optimal probability distributions. However, the learning results are smaller. This is because the secondary users in the spectrum access game are selfish and only want to maximize their own utility. Thus, although they agree to simultaneously increase their access probabilities gradually, once a secondary user can not getter higher utility by increasing its access probability, the learning algorithm is forced to stop. While the
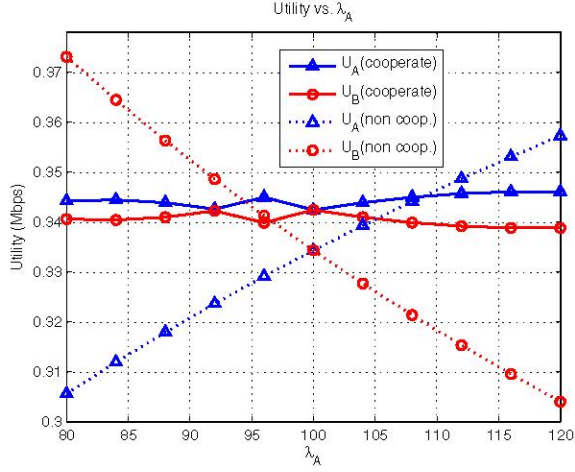
Fig. 5: Comparison of utilities under cooperation and noncooperation.

Fig. 6: Utility sum ($U_A + U_B$) under cooperation and noncooperation.

theoretical PF optimal probabilities are obtained without considering the selfishness of the secondary users. Therefore, the learned access probabilities are smaller than the theoretical optimal ones.

We can also observe from Figure 4 that the optimal access probability for user B is almost the same, regardless of λ_A, and user A's access probability is reduced when λ_A is increasing. This is because from the simulation parameters, we have $r_1^B = r_1^A > r_2^B = r_2^A$, so the actual admitted access request rates $(a_\gamma \lambda_\gamma)$ for user A and B should be the same to ensure fairness. This also shows that the self-learning repeated game is fair in the sense that the secondary user's access probability is only determined by its own traffic statistics and the channel conditions, and not by other opponent user's traffic patterns.

In Figure 5, we compare the utility values versus λ_A with cooperation and noncooperation of the secondary users. As is mentioned above, since $r_1^B = r_1^A > r_2^B = r_2^A$, theoretically the two secondary users should be given equal chance $(a_\gamma \lambda_\gamma)$ to access the licensed band no matter how λ_A changes. With the proposed self-learning repeated game, we can see the secondary users almost have the same utilities, so fairness is ensured. However, when neither of the secondary users cooperates and accesses the spectrum band with probability 1, i.e., plays the NE of the static game, the utility distribution for the secondary users is very unfair. In Figure 6, we compare the sum of the utility values versus λ_A with cooperation and noncooperation of the secondary users. We observe that with cooperation the secondary users can obtain a larger aggregate utility than that of
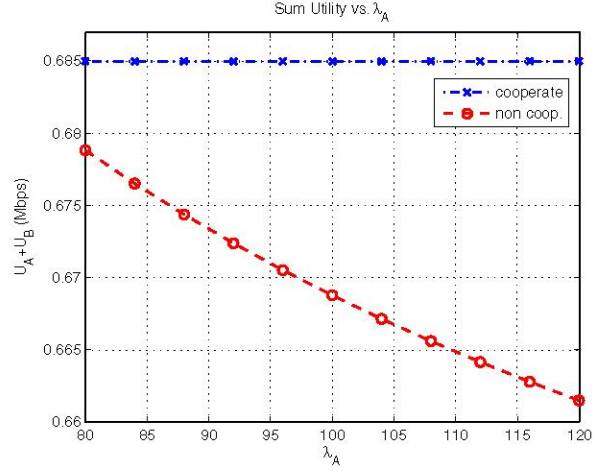
the noncooperation case. As the traffic injected by the secondary users becomes heavier, the difference of the utility sum between cooperation and noncooperation becomes even greater. This fact indicates that cooperation among secondary users is especially beneficial when there is heavy traffic in a dynamic spectrum access network. If more secondary users share the spectrum and suffer more severe mutual interference, we can also expect that by cooperation, the secondary users can achieve much better aggregate utility than that of the noncooperation case.

V. CONCLUSIONS

In this paper, we propose a self-learning repeated game framework for distributed primary-prioritized dynamic spectrum access. In order to coordinate secondary users' access adapt to the spectrum dynamics, the interactions among secondary users are modeled as a repeated game. To improve the spectrum efficiency under the selfishness of secondary users, we develop a self-learning algorithm that helps secondary users obtain their optimal access probabilities with only local observations. The simulation results show that the proposed self-learning algorithm has fast convergence, enhances the game efficiency, and achieves comparable performance with those of the PF dynamic spectrum access.

REFERENCES

[1] FCC, "Spectrum policy task force report," *FCC Document ET Docket No. 02-135*, Nov. 2002.
[2] FCC, "Facilitating opportunities for flexible, efficient and reliable spectrum use employing cognitive radio technologies: notic eof proposed rule making and order," *FCC Document ET Docket No. 03-108*, Dec. 2003.

[3] J. Mitola III, "Cognitive radio: an integrated agent architecture for software defined radio," Ph.D. Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.

[4] Y. Xing, R. Chandramouli, S. Mangold, and S. Shankar N, "Dynamic spectrum access in open spectrum wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 626-637, Mar. 2006.

[5] S. Keshavamurthy and K. Chandra, "Multiplexing analysis for spectrum sharing," in *Proc. of IEEE MILCOMM'06*, Washington, D.C., Oct. 2006.

[6] B. Wang, Z. Ji, and K. J. R. Liu, "Primary-prioritized Markov approach for dynamic spectrum access," to appear in *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'07)*, 2007.

[7] L. Cao and H. Zheng, 'Distributed spectrum allocation via local bargaining," in *Proc. of IEEE SECON 2005*, pp. 475-486, Santa Clara, CA, Sep. 2005.

[8] J. Huang, R. A. Berry, and M. L. Honig, "Auction-based Spectrum Sharing," in *ACM/Springer Mobile Networks and Applications Journal (MONET)*, vol. 11, no. 3, pp. 405-418, Jun. 2006.

[9] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," in *Proc. of IEEE DySPAN 2005*, pp. 251-258, Baltimore, MD, Nov. 2005.

[10] Z. Ji and K. J. R. Liu, "Belief-assisted pricing for dynamic spectrum allocation in wireless networks with selfish users," in *Proc. of IEEE SECON 2006*, pp. 119-127, Reston, VA, Sep. 2006.

[11] Z. Ji and K. J. R. Liu, "Collusion-resistant dynamic spectrum allocation for wireless networks via pricing," to appear in *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'07)*, 2007.

[12] Z. Ji and K. J. R. Liu, "Dynamic spectrum sharing: a game theoretical overview," to appear in *IEEE Communications Magazine* (Feature Topic on Cognitive Radios for Dynamic Spectrum Access), May 2007.

[13] IEEE 802.22 working group on wireless regional area networks, http://www.ieee802.org/22/.

[14] —, "Improving the sharing of the radio spectrum: final report", http://www.ofcom.org.uk/research/technology/overview/ese/share/.

[15] FCC, "Establishement of interference temperature metric to quantify and manage interference and to expand available unlicensed operation in certain fixed mobile and satellite frequency bands," ET Docket 03-289, Notice of Inquiry and Proposed Rulemaking, 2003.

[16] T. Clancy, "Formalizing the interference temperature model," to appear, *Wireless Communications and Mobile Computing*.

[17] G. Owen, *Game Theory*, 3rd ed. New York: Academic, 2001.

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