Distributed Learning in Secondary Spectrum Sharing Graphical Game

Mahdi Azarafrooz and R. Chandramouli
Depatment of Electrical and Computer Engineering
Stevens Institute of Technology

***Abstract*—Secondary users sharing primary users' spectrum is modeled as a graphical game. Users located in random graphs and a regular lattice are considered. Secondary users are assumed to differentiate the "quality" of the primary spectrum while interacting within their local neighborhood to minimize interference and congestion. The learning algorithm is also shown to be effective in punishing malicious users that violate spectrum etiquettes. An equivalence between spectrum sharing neighborhood interaction and the spin-glass model in statistical physics is established. A distributed exponential learning algorithm is used to arrive at an evolutionary stable solution to the game. Some theoretical properties of the system are studied and simulation results are presented to illustrate price of anarchy, convergence of the learning algorithm and asymptotic invariance of the system performance with respect to spectrum quality.**

***Index Terms*—Graphical Game Theory, Local Minority Game, Evolutionary Stable strategy, Spin Glass model, Partition function, Potential function, Spectrum Selection**

I. Introduction

Cognitive radio (CR) nodes learn to configure their transmission and reception parameters based on measurements both external and internal to the radio environment, to achieve efficiencies in spectrum utilization and related metrics. We address the distributed optimization of secondary user sharing of primary user spectrum considering the spatial re-use as well. This is modeled as a spatial or graphical game theoretic problem where radio interference and congestion induced by local neighbor communication is taken into account. Congestion is defined as the number of secondary users communicating in a specific band. An unsupervised local learning algorithm is investigated to solve the game problem. A connection between this problem and the spin glass problem in statistical physics is established. Several theoretical results are presented along with numerical simulations.

In [4] spectrum sharing and spatial reuse in a wireless network is posed as an extended form of the congestion game where users' payoff for using a spectrum band or channel is a function of the number of its interfering users sharing that channel. In [5] spectrum management is studied in CR by defining a secondary user specific utility as a function of spectrum opportunity, congestion and bandwidth.

Minimizing congestion in spectrum sharing lends itself naturally to be modeled as a minority game (MG) [1]. That is, if a minority of the users choose to communicate in a band then they all receive a higher payoff due to a smaller level of congestion. It has been noted (e.g., [10]) that deductive reasoning methods to solve the MG problem may fail, for example, when the players have bounded rationality. In a complex wireless network, when a CR node interacts with other CR or non-CR nodes it is unreasonable to assume that the nodes can use deductive reasoning to compute their optimal strategies. Therefore, to address this issue in the spatial minority spectrum sharing game we explore an inductive scheme. The behavior of selfish nodes that dynamically switch their channels using broadcasted random public signal is presented in [2]. In [3] dynamic spectrum access is modeled as a MG where the networks try to minimize their cost in finding a clear band. A graphical game model for competitive spectrum access is discussed in [11]. A fundamental question is: can local inductive approaches such as simple rules of thumb solve the spectrum sharing game (global) optimally? We answer this question in this paper.

II. Model and Analysis

We consider a heterogeneous CR network scenario with an arbitrary network topology, having n secondary users with access to B primary user bands. Heterogeneity implies that the secondary users may employ different network protocols having different requirements, e.g tolerance for delay, transmission rate, etc. The secondary users compete for spectrum opportunities in a decentralized non cooperative manner. Let k_i denote the index of spectrum band that user i is active in and $\mathcal{X}_{k_i}$ be the number of active users in the band k_i and $\mathcal{I}_i^{k_i}$ is the number of transmissions interfering with user i in band k_i. Let y_l, $l = \{1, 2, \ldots, B\}$ denote the quality of the l^{th} spectral band. For example, this could be the bandwidth, capacity, data

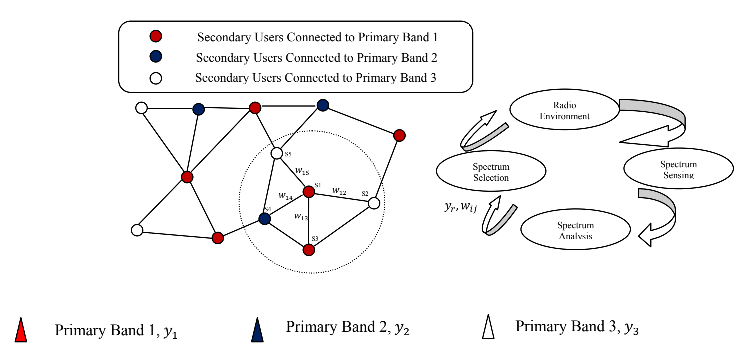


Fig. 1. Secondary spectrum sharing graphical game example.

rate, etc. We assume $\{y_l\}$ to be public knowledge to all the secondary users and higher the value the more desirable that band is. Without loss of generality let $\sum_{i=1}^{B} y_l = 1$ where $0 \leq y_l \leq 1, \forall l$. The utility obtained by secondary user i is $U_i(\mathcal{X}_{k_i}, \mathcal{I}_i^{k_i}, y_{k_i})$. That is, the utility function depends on the interference level, congestion level as well as the quality of the operating band. Then the optimal solution to the spectrum sharing problem is given by (1):

$$\{\overline{K} = \underset{k_i}{\operatorname{argmax}} \sum_i U_i(\mathcal{X}_{k_i}, \mathcal{I}_i^{k_i}, y_{k_i}), \quad \forall i \in \{1, ..., n\}, \forall k_i \in \{1, ..., B\}\} \tag{1}$$

The graphical spectrum selection game (GSSG) as depicted in Fig. 1 is defined as:

1. $(\mathcal{G}, \mathcal{M})$ where each user i is represented by a vertex in an undirected graph $\mathcal{G}$ and $\mathcal{M}$ is the set of n local game matrices M_i. Note that M_i depends on $U_i(.)$. Let $\mathcal{N}(i) \subseteq \{1, ..., n\}$ denote the neighborhood of user i in $\mathcal{G}$, that is, those vertexes j such that the edge (i, j) appears in $\mathcal{G}$.
2. Players are the secondary users $i \in \{1, ..., n\}$.
3. Set of pure strategies for user (vertex) i is the set of $S_i = \{1, ..., B\}$. The mixed strategies for user i is the probability mass function $p_i(s) \equiv p_{is}, \sum_{s=1}^{B} p_{is} = 1$ on S_i. Also let k_i denote the strategy chosen (state) by a specific vertex i. Let $k : \mathcal{G} \rightarrow \mathcal{S}$ where $S \in \{1, ..B\}^n$ denotes the joint strategy of all the n secondary users.
4. Then the utility user i receives, v_i, is given by:

$$v_i(k_i) = \sum_{j \in \mathcal{N}(i)} w_{ij} M_i(k_i, k_j) \tag{2}$$

where w_{ij} is the weight of interaction between player i and j (Fig. 1).

Then, the optimal strategy for a selfish user i that maximizes its payoff is given by:

$$\overline{k_i} = \underset{k_i \in \mathcal{S}}{\operatorname{argmax}} \, v_i(k_i) \tag{3}$$

Now for example consider a two-user game where $M_1 = M_2 = ...M_n = M$ where M is the following 2×2 anti-coordination game matrix:

$$\begin{pmatrix} 0,0 & y_1, y_2 & ... & y_1, y_B \\ y_2, y_1 & 0,0 & ... & y_2, y_B \\ ... & ... & ... & ... \\ y_B, y_1 & y_B, y_2 & ... & 0,0 \end{pmatrix}$$

It is clear from this matrix that if both the secondary users choose the same band then they both receive zero pay-off.

It should be noted that although the payoff to player i are determined only by the actions of the players in $\mathcal{N}(i)$, equilibrium still requires global coordination across the player population. We will show that secondary users playing anti-coordination games with the neighbors using an exponential learning algorithm, reduces local interference in a neighborhood and the resulting evolutionary stable strategy provides a fair and optimal solution.

It is easy to show that for the anti-coordination game M described above the potential function exists. For example in the homogeneous case ($y_i = 1/B$, $\forall$ i) the potential function H is:

$$\begin{pmatrix} -1 & 0 & ... & 0 \\ 0 & -1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & -1 \end{pmatrix}$$

Note that not every anti-coordination game has a potential function.

Theorem 1. If an anti-coordination game matrix M has potential function H, then so does the associated "graph game" $(\mathcal{G}, \mathcal{M})$ on any weighted graph with the potential function:

$$\mathcal{H}(k) = \sum_i \sum_{j \in \mathcal{N}(i)} w_{ij} H(k_i, k_j) \tag{4}$$

where $\mathcal{H}$ is the potential function for graphical $(\mathcal{G}, \mathcal{M})$.

Proof. To see this suppose that agent i deviates, say by choosing $k_{i'}$. Let $k' = (k_{i'}, k_{-i})$. Then,

$$\begin{aligned} v_i(k) - v_i(k') &= \sum_{j \in \mathcal{N}(i)} w_{ij} [M(k_i, k_j) - M(k_{i'}, k_j)] \\ &= \sum_{j \in \mathcal{N}(i)} w_{ij} [H(k_i, k_j) - H(k_{i'}, k_j)] \end{aligned}$$

It follows:

$$\mathcal{H}(k) = \sum_i \sum_{j \in \mathcal{N}(i)} w_{ij} H(k_i, k_j)$$

For simplification lets $w_{ij} = 1$ up to section III.

Lemma 1. Maximizing the potential function $\mathcal{H}$ of the defined graphical game is isomorphic to minimizing the energy function of a " antiferromagnetic Spin-Glass Model" E .

In statistical physics the energy function of Spin-Glass model [7] is defined to be $E = \sum_{i \neq j} \mathcal{J}_{ij} \sigma_i \sigma_j$ where σ_i represents the spin of the particles and coupling coefficients $\mathcal{J}_{ij} \geq 0$ are limited to neighborhood interactions. In a 2-spin model the spin value σ_i is +1 or -1. Now if the the spin values of two neighboring particles are the different then the value of the energy function will decrease. Similarly it can be easily seen from (4) that if two neighboring secondary nodes transmit in different channels the potential function will increase. Then by interpreting different spins are different spectrum bands and by considering a neighborhood limited range of interactions among the particles (secondary users) one can easily prove the equivalence between energy function and potential function of proposed game. We use this lemma later in the proof of Theorem 2.

A. *Nash Equilibrium and Price of Anarchy*

To find the Nash equilibrium of the defined graphical game we have to compute the optimal configuration $\overline{k}$ that maximizes the potential function or minimize the energy of the Spin-glass model. The optimal configuration gives the ground states of the system. Ground state is the lowest-energy state

of the system. But computing ground states of a spin glass model is a NP-complete problem [7] in general.
The price of anarchy (PoA) is the ratio between the maximum attainable payoff and the one attained by the game's equilibrium solution.

B. Learning and Evolutionary Stable Strategy (ESS)

Since the defined game $(\mathcal{G}, \mathcal{M})$ is a minority game it can be solved using an inductive reasoning or learning method rather than a deductive one as discussed earlier. Exponential learning (see, (5) & (6)) is an inductive manner since it does not determine which beliefs are a priori rational for a user, but rather determines how the beliefs should be updated in a deductive manner from the evidence obtained from environment. Each secondary user i starts with an initial probability Y of choosing one of the B available strategies. Then for the chosen strategy obtains the value/reward $v_i(s)$. The cumulative reward is kept

Algorithm 1 Local Exponential Learning Algorithm

Initialization: $t = 1$, every secondary user chooses spectral band according to vector $Y = [y_1, ..., y_B]$.
1. Every node i broadcast the chosen spectral band to its neighbors $\mathcal{N}_i$.
2. Users evaluate utility $v_i(s,t)$ using (2) $\forall s \in \{1, ..., B\}$.
3. Users update their mixed strategy profile at iteration t $P_i(t) = [p_{i1}(t), ..., p_{iB}(t)]$ according to (5) & (6).
4. Users select spectral band using mixed strategy profile $P_i(t)$ for iteration $t = t + 1$.
5. Stop if:
$\max_s |v_i(s,t) - v_i(s,t-1)| \leq \beta, \beta \geq 0 \;\; \forall i \in \{1, ..., n\}$.
Otherwise go back to step 1.

track of and the probability of choosing strategy s at time t is updated according to the following equations:

$$\mathcal{U}_{is}(t+1) = \mathcal{U}_{is}(t) + v_i(s,t) \tag{5}$$

$$p_{is}(t) = \frac{e^{\Gamma \mathcal{U}_{is}(t)}}{\sum_{s'=1}^{s'=B} e^{\Gamma \mathcal{U}_{is'}(t)}}, \forall s \in \{1, ..., B\} \tag{6}$$

where Γ is the learning coefficient. The steps involved in the spectrum selection process is described in Algorithm 1.
Proposition 1 [15, 16]. $\mathcal{U}_i(t)$ is the sum of the achieved payoff for user i up to iteration t, $\delta \in (0 \;\; 1)$, then the proposed learning algorithm (5) & (6), with $\Gamma = L^{-1}\sqrt{\frac{8|\mathcal{N}_i|}{t}}$ satisfies, with probability at least $1 - \delta$,

$$[\max_{k_i \in \mathcal{S}} \mathcal{U}_{ik_i}(t) - \mathcal{U}_i(t)] \leq \sqrt{\frac{t}{2} \ln |\mathcal{N}_i|} + \sqrt{\frac{t}{2} \ln (\frac{1}{\delta})}, \quad \forall i \in \{1, ..., n\} \tag{7}$$

where $|\mathcal{N}_i|$ is number of user i's neighbors and $L = \max\,[y_1, y_2 ... y_B]$.
Proposition 1 indicates that the average regret:

$$[\max_{k_i \in \mathcal{S}} \mathcal{U}_{ik_i}(t) - \mathcal{U}_i(t)]/t, \forall i \in \{1, ..., n\}$$

converges to zero with the rate of $|\mathcal{N}_i|^{-1/2}$.
Theorem 2. The price of anarchy in the defined graphical game does not depend on $Y = [y_1, y_2, \ldots, y_B]$ in the steady state.
Proof. The proof depends on the characteristics of spin glass model and is not given here due to space constraints.
Theorem 3. Nash Equilibrium of any anti-coordination game is evolutionary stable under exponential learning dynamics [12].
Theorem 4 [9]. Let $\mathcal{G} = \{\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)\}_{n=0}^{\infty}$ be an infinite family of graphs in which for all $v \in \mathcal{V}_n$, $deg(v) = \Omega(n^\gamma)$ (for any constant $\gamma > 0$). Let M be any 2-player, symmetric game, and suppose s is a classical ESS of M. Let t be any mutant strategy, and let the mutant family $\mathcal{F} = \{\mathcal{F}_n\}_{n=0}^{\infty}$ be chosen randomly by labeling each vertex a mutant with constant probability ϵ, where $\epsilon_t > \epsilon > 0$. Then with probability 1, s is an ESS with respect to M, $\mathcal{G}$ and $\mathcal{F}$.
Theorem 3 and 4 guarantees that the $s = [y_1, y_2 ... y_B]$ as the Nash Equilibrium of the proposed anti-coordination game to be the evolutionary stable strategy of spectrum sharing graphical game. As a result of this, in the evolutionary stable state of the system each channel i is chosen by the y_i fraction of population, i.e., proportional to the spectrum quality of primary bands.

Now, if:

$$U_i(\mathcal{X}_{k_i}, \mathcal{I}_i^{k_i}, y_{k_i}) = \mathcal{X}_{k_i} y_{k_i}/(1 + \mathcal{I}_i^{k_i}) \tag{8}$$

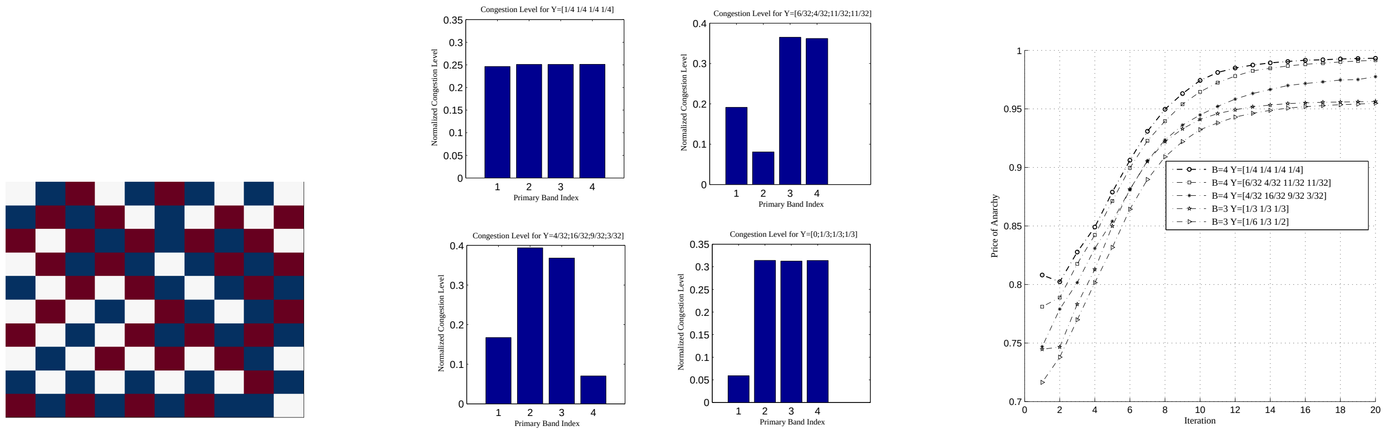
it can be easily seen that if user i shares sharing the band k_i with the quality measure y_{k_i} it also suffers from a congestion level of $1/y_{k_i}$ as a result of the evolutionary stable solution strategy of the system which gives $\mathcal{X}_{k_i} = 1/y_{k_i}$. This results in the payoff equal to $1/(1 + \mathcal{I}_i^{k_i})$. However the interference from the local neighbors will be zero if the number of primary bands are large enough.

III. Punishing Spectrum Etiquette Violation

CR nodes not only learn from their personal observation from the radio environment but also interact with other nodes (in our case neighbors). Therefor it is possible for nodes to free ride on the others or maliciously cheat the system to increase their chance of spectrum access through etiquette violations. The idea of punishing nodes can be by setting the weights $w_{ij} \in [-\infty \;\; 1], j \in \mathcal{N}_i$ in the graphical game. For example if $w_{ij} = 0$, it means that the secondary node i ignores the outcome of interaction with user j. If $w_{ij} = -\infty, \forall j \in \mathcal{N}_i$ then all the neighbors punish node i by jamming it, i.e; they transmit in the same band as i w.p. 1. This mimics the 'Homo Reciprocans' behavior as discussed in [14].

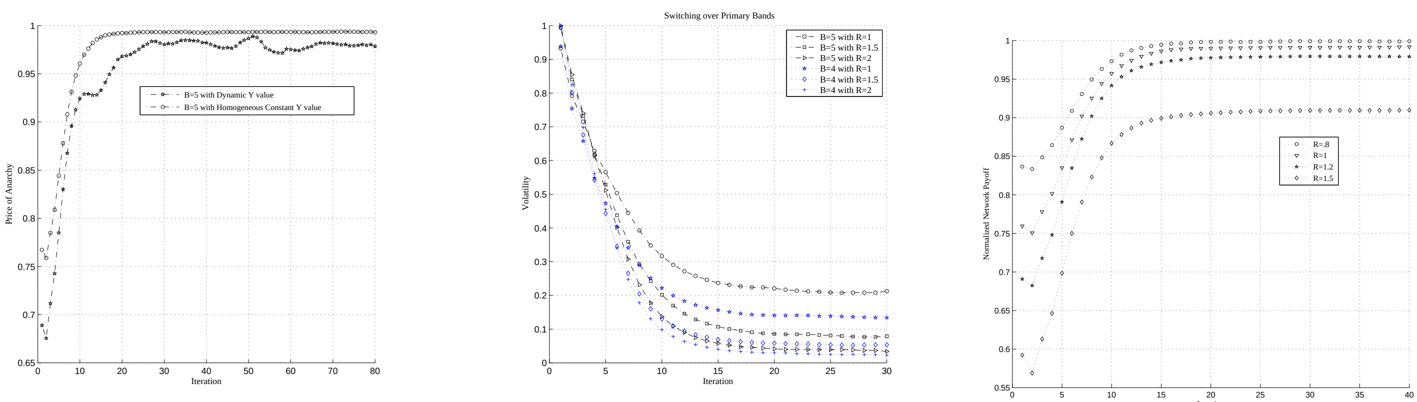
IV. Simulation Results

We assume a square area $A = 100$ square units uniformly distributed random configuration of $n = 100$ secondary nodes. We consider different communication ranges, R in the simulations. Specifically, we choose $R = \{0.8, 1, 1.2, 1.5, 2\}$. That is, $R = 0.8$ means that for a secondary node all other nodes within an Euclidean distance of 0.8 are considered to



a. *b.* *c.*

Fig. 2. a) Spatial configuration of secondary users spectrum selection for $B = 3$ denoted by three different colors. b) Congestion levels of the primary bands for different spectrum quality vector Y validating the result of theorems 3 and 4; R=1 averaged over 100 realizations of the game. c) The price of anarchy: As predicted by Theorem 2, different spectrum qualities do not affect the price of anarchy; R=.8 averaged over 100 realizations of the game.



a. *b.* *c.*

Fig. 3. a) Price of anarchy when the Y-vector changes dynamically every 8 iteration, R=1.2 averaged over 100 realizations. b) Decreasing R with fixed number of strategies leads to more channel switching similar to that of minority game, there should be a balance between the number of strategies and average connectivity of the graph; averaged over 50 realizations of the game. c) Increasing the range of interaction degrades the performance of the system, showing the trade-off between switching cost and normalized payoff of the network; averaged over 100 realizations of the game.

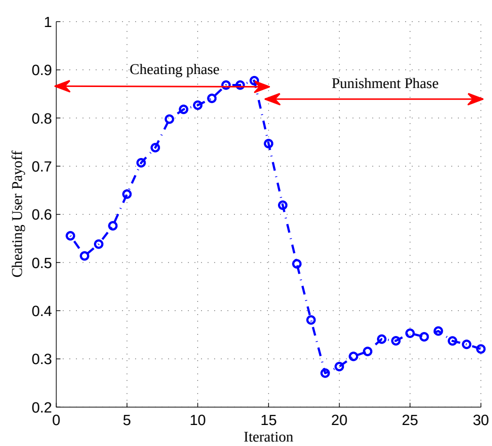


Fig. 4. Learning to punish a cheating user with three neighbors. Starting in iteration 15, the neighbors converge on the strategy by interfering in its band; $R = 1.2$, $B = 4$, averaged over 50 realizations of the game.

be its neighbors. We chose $\Gamma = 1$ and different values for $Y = \{y_1, y_2, \dots, y_B\}$. Fig. 2a is the Nash equilibrium solution obtained using exponential learning, to the spatial spectrum sharing game that minimizes the interference when $B = 3$ and $R = 1$ in a lattice configuration. We observe from this figure that majority of the neighbors (up, down, left, right) of a secondary user receive a different spectrum allocation. There are only few cases where this is not the case. From Fig. 2b we see that the evolutionary stable strategy results in congestion levels that are proportional to the values y_i, $\forall i$. That is, the learning algorithm allocates spectrum bands to secondary nodes proportional to the bands' quality. Fig.2c shows that the price of anarchy does not depend on Y given by Theorem 2. The primary spectrum quality vector Y may not be a constant always. For example when the r^{th} primary user starts transmitting then the corresponding value y_r will become smaller. Fig. 3a shows that even in the case of dynamically changing Y-vector the price of anarchy asymptotically becomes independent of Y, if Y changes relatively slowly compared to the convergence of the learning algorithm.

We define switching cost as follows. If a user switches from one band to another during a learning iteration then the switching cost is equal 1 otherwise it is 0. Then average switching cost. The volatility is the variance of the switching cost computed over the period of the iterations. The normalized network payoff is the average payoff for the entire network. We see from Fig. 3b that a lower value of R for a fixed B incurs higher volatility. This is because a lower value for R means a smaller neighborhood range and therefore the secondary users explore several configurations (choices of primary bands) due to the stringent constraint in order to minimize interference and congestion. But as the learning converges the volatility decreases indicating convergence towards an optimal configuration of spectrum bands for the secondary users.

Fig. 3c shows the effect of increasing the interaction range R for a fixed B on the normalized network payoff. It is clear from the figure that as R increases the normalized network payoff value decreases. This is due to the fact that a higher R value results in a larger neighborhood with more number of secondary nodes in the neighborhood, on an average. Therefore, as the learning algorithm converges the shared spectrum results in a higher interference leading to a lower network payoff.

Fig. 4 shows that a malicious user is able to cheat the system via spectrum etiquette violations for up to 15 iterations. During this phase the cheating user is able to increase its payoff. But, the other interacting secondary users in its neighborhood are able to learn this behavior and adjust their mixed strategies to punish. We see from the figure that the payoff of the cheating user starts decreasing during the punishment phase.

V. Conclusion

We showed that there is an equivalence between the secondary spectrum sharing graph game and the spin-glass model. A distributed exponential learning algorithm results in an evolutionary stable solution for the game. There is an inherent relationship between the number of available primary bands, communication range of the secondary nodes that defines its neighborhood and the volatility in spectrum switching cost. The price of anarchy is independent of the primary user spectrum quality (asymptotically) for both the static and dynamic cases. The congestion levels in the spectrum band computed by the learning algorithm is proportional to the quality of the spectrum bands. The distributed learning algorithm enables the interacting neighbors to detect and punish a malicious or cheating user that violates spectrum etiquettes to maximize its payoff.

References

[1] M. Marsili, D. Challet and R. Zecchina, "Exact solution of a modified El Farols bar problem: Efficiency and the role of market impact,"*Physica A*, vol. 280, pp. 522 553, 2000.

[2] P. Mertikopoulos and A. Moustakas, "Correlated Anarchy in Overlapping Wireless Networks",*IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1160 - 1169, 2008.

[3] S. Sengupta, R. Chandramouli, S. Brahma and M. Chatterjee, "A game theoretic framework for distributed self-coexistence among IEEE 802.22 networks ", *IEEE GLOBECOM*, pp. 1-6, 2008.

[4] M. Liu and Y. Wu, " Spectum sharing as congestion games", *Annual Allerton Conference on Communication, Control, and Computing*, pp. 1146 - 1153, 2008.

[5] I. Malanchini,M. Cesana and N. Gatti, " On Spectrum Selection Games in Cognitive Radio Networks", *IEEE GLOBECOM*, pp. 1 - 7, 2009.

[6] H. Peyton Young, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, 1st ed. Princeton University Press, 2001.

[7] S. Kirkpatrick and D. Sherrington, "Infinite ranged models of spin-glasses", *Physical Review B*, vol. 17, no. 11, 1978.

[8] J. Hofbauer," Stability for the best response dynamics", mimeo. 1995.

[9] M. Kearns and S. Suri, "Networks preserving evolutionary equilibria and the power of randomization". *In Proceedings of the 7th ACM Conference on Electronic Commerce*, 2006.

[10] W.B. Arthur, "Inductive reasoning and bounded rationality (the El Farol problem)". *Am. Econ. Assoc. Papers Proc., vol. 84, pp. 406411*, 1994.

[11] H. Li, Z. Han, "Competitive Spectrum Access in Cognitive Radio Networks: Graphical Game and Learning". *Wireless Communications and Networking Conference (WCNC)*, 2010.

[12] F. Kojima, S. Takahashi, "Anti-Coordination Games and Dynamic Stability". *International Game Theory Review (IGTR), vol. 09, issue 04, pages 667-688*, 2007.

[13] S. Moelbert, P. Rios, "The Local Minority Game". *Physica A, 303(1-2), 217-225*, 2001.

[14] Y. Xing, R. Chandramouli, "Human behavior inspired cognitive radio network design "*IEEE Communications Magazine*, vol. 46, pp. 122-127, 2008.

[15] G. Stoltz, R. Lugosi, "Internal Regret in On-line Portfolio Selection "*Mach. Learn*, vol. 59, pp. 125, 2005.

[16] N. Cesa-Bianchi, G. Lugosi, *Prediction, Learning and Games*, Cambridge University Press, 2006.