Self-Coexistence in Cognitive Radio Networks using Multi-Stage Perception Learning

Deepak K Tosh
Department of Computer Science
Graduate Center, CUNY, NY
Email: dtosh@gc.cuny.edu

Shamik Sengupta
Department of Comp. Sc. & Engr.
University of Nevada, Reno
Email: ssengupta@unr.edu

***Abstract*—In this paper, we study the self-coexistence problem among competitive Cognitive Radio (CR) networks in an uncoordinated distributed wireless environment of homogeneous and heterogeneous bands. This problem can be correlated with famous optimal foraging theory, where the humming birds forage to explore islands in search of food sources to survive. The behavior of learning from observations leads them to find island of optimal resources. The proposed perception based learning mechanism for homogeneous spectra, helps the CR networks to strategize their choice of actions on the basis of rewards gathered from the accessed spectrum bands and successfully grab a clear chunk of spectrum. However, in heterogeneous bands scenario, the CR networks inadvertently choose the best suitable band greedily which lead them to collision. We incorporate a regret minimization technique with the proposed learning mechanism to resolve the contention among them and maximize system performance. Experimental results conclude that the networks could achieve the objective of finding a free spectrum with maximized system utility using the proposed heuristic within limited number of interactions.**

I. Introduction

Over the past decade, wireless services and applications are exponentially expanding their horizon along with population of users. However, the survey conducted by Federal Communications Commission (FCC) found spatial wastefulness of spectrum resources due to traditional fixed allocation scheme and proposed new policy regulation that allows dynamic spectrum access (DSA) of unused spectra opportunistically in a non-interfering basis. This functionality can be best exploited by the cognitive radios (CR) which are supposed to learn about availability of spectrum bands by periodic sensing so that spectrum holes can be reused for data communication by tuning its transmission parameters without interfering with primary incumbents transmission.

In a distributed uncoordinated wireless environment, the CR networks fight for spectrum resources aiming to minimize interference with neighboring networks which will improve the probability of successful transmission and thereby performance. To coordinate each rational CR networks in finding best available spectrum band in this bazaar environment is indeed challenging. Thus strategic thinking among CR agents can help them to coordinate and maintain self-coexistence. Past works on CR have been concentrating on primary–secondary spectrum etiquettes, spectrum sensing, primary user detection techniques etc., but not much work have been done on the self-coexistence issues, and more importantly self-coexistence in heterogeneous spectrum scenario. To maintain self-coexistence among IEEE 802.22 base stations (BS), a utility graph coloring technique is proposed for allocating spectrum to BSs in [3]. In [4], the distributed self-coexistence with homogeneous bands is modeled as modified minority game and solved to find mixed strategy Nash equilibrium for the game. A Coexistence-Aware Spectrum Sharing (CASS) protocol was proposed in [6] which minimizes the self-interference with minimum control overhead. To improve spectrum utilization, a round-robin based resource allocation algorithm was proposed for IEEE 802.22 WRAN in [7] which maintains the fairness of resource allocation to improve spectrum utilization.

In this paper, the self-coexistence among CR networks is studied in a distributed wireless environment where CR networks compete to find a contention free spectrum band that maximize their overall reward over a long run. This competition model is similar to the famous optimal foraging model [11], where a species of birds forage over different islands to find food sources of good amount by investing their energy in flying. The basic trade-off lies between total energy gain from the food sources and scavenging period for long survival. As there might be more species scavenging for food in the same island, the bird's foraging period can be affected with this contention too. Here the network players (birds) forage for spectrum bands (islands) in an uncoordinated manner to maximize their throughput. Here the networks need to adapt by learning from its own actions and payoffs, so that its overall gain is maximized. When spectrum resources are categorized as homogeneous or heterogeneous types, the behavior of CR networks is also studied here.

The contributions in the paper are as follows: (1) For homogeneous bands, we present a perception based learning model which helps networks in building perception about spectrum bands by observing the payoffs. And the perception vector is mapped to find strategy of deciding whether to explore or exploit. (2) For heterogeneous bands, the networks aim to achieve two goals simultaneously: get a contention free band and the average utility reward over the stages must be maximized in the non-cooperative simultaneous move game.

† "This research has been funded by NSF grant CNS # 1149920."

978-1-4673-6187-3/13/$31.00 ©2013 IEEE

No signaling is allowed between the players[1], thus they should learn by observation and use the previous history to achieve both goals. The proposed regret minimization heuristic helps to achieve optimal system utility in this heterogeneous scenario.

The rest of the paper is organized as follows. The system assumptions and description of self-coexistence among CR networks competing for homogeneous bands are elaborated in the section 2. We also present the perception based learning model in this section. In section 3, we study the heterogeneous self-coexistence problem and propose a regret minimization based heuristic to achieve optimal system utility. The experimental details and results of the conducted simulations are analyzed in section 4. Finally concluding remarks are presented in the last section.

II. Self-Coexistence in Homogeneous Bands

A. System Description

In this work, the self-coexistence problem is modeled as a dynamic multi-stage interaction game where N CR networks compete for accessing M distinct orthogonal spectrum bands. The rational CR networks are homogeneous in nature with same strategy space. They aim to access exclusively one out of M spectrum bands that are free of contention from primary users. This foraging game is studied based on homogeneous and heterogeneous resources, because when islands are homogeneous and provide identical amount of food source, rational birds will try to find any island where no other entity is scavenging, however when islands provide non-identical amount of food source, birds will optimally forage to find the best possible island which will tend to satiate the bird's need. In our case, the band heterogeneity depend on factors like bandwidth, data rate, operating frequency range etc. So self-coexistence in heterogeneous bands will be interesting to analyze because the rational player's foraging behavior will be different from homogeneous case. In this spectrum foraging game where spectrum bands are equal in characteristics, networks always compete to grab a contention-free channel as quick as possible. In the following subsections, we formally describe the game settings for homogeneous band based self-coexistence problem and a perception based learning model for all networks to optimally scavenge to acquire a band.

B. Game Settings for homogeneous band based self-coexistence problem

For the homogeneous scenario, it is considered that each network tries to acquire a spectrum band within minimum scavenging period which will be used exclusively by one CR network. Accessing a particular band by multiple networks results in no reward to each of the contending networks. Each network i has a mixed strategy space to play in this game: $p_i = (p_1^{(i)}, p_2^{(i)}, ..., p_M^{(i)})$, where $0 \leq p_j^{(i)} \leq 1$ is the probability of network i choosing the band j to operate and $\sum_{j=1}^{M} p_j^{(i)} = 1$. Because the bands are assumed to be homogeneous in nature, therefore returns a constant utility on each access. Thus the utility function for player i can be defined as follows:

$$U_i(a_i, a_{-i}) = \begin{cases} \alpha & \text{if } a \neq a_i,\ \forall a \in a_{-i} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where α is a constant utility for all spectrum bands.
According to above utility function, if a network i chooses band j to operate, then it obtains a constant utility of α provided no other network has chosen the same band j. If two or more networks choose the same band j out of M bands, then all contending networks receive no reward, rather there exist some cost for initiating communication and sensing the band for availability. In the foraging process, it is assumed that the birds do not have any information on actions or strategies of other birds while deciding which island to forage. Similarly each network takes independent action based on its perception about spectrum bands. The decision of a network at the end of a stage can be either to stick with currently chosen band and exploit it, if the network player is satisfied with the current gain, or choose another band to explore more. The following described perception based learning model helps to strategize the player's action based on their accumulated perception vector.

C. Perception based Learning Model

This learning model helps the players to build belief/perception about the accessed spectrum bands based on perceived utility. The perception of a network can be interpreted as a metric for classifying the spectrum bands based on its throughput reward. By statistical analysis from the observed data, the players try to minimize the probability of interference in accessing a spectrum band. Each network $i \in N$ maintains a perception vector, $P^{(i)} = (P_1^{(i)}, P_2^{(i)}, P_3^{(i)}, ..., P_M^{(i)})$, about all M bands, and updates j^{th} entry after accessing to band j. The perception $P_j^{(i)}$ of a player i about a band j is mapped to player i's mixed strategy. Based on the generated strategy, the network takes stochastic decision about choosing any band $j \in M$, for next game stage. Based on observed reward from band j by network i, perception value band j, $P_j^{(i)}$ is updated. All networks aim to get a band free of contention for the purpose of co-existence with less number of game interactions.

In the starting of foraging stage $t = 0$, network $i \in N$ chooses a band $j \in M$ randomly to operate. The observed utility in stage t is defined to be $U_{i,a_i(t)}(t)$. As there exist no history about game until stage $t = 0$, we set the initial perception vector of player i $(P^{(i)}(0))$ to a small constant to avoid biasness towards any particular band. In many past literatures on reinforcement learning[9][10], the Q-parameters for actions are estimated from the player's experience. And the Q-values are mapped to the player's mixed strategy based on Boltzmann distribution which is a common softmax method for controlling the exploration in a large search space using a controlling parameter named temperature $(\gamma > 0)$. In our model, the perception vector is obtained according to the player's experience on actions. Thus the perception about

[1]We use the words "player", "network" interchangeably throughout the paper.

choosing action in next stage can be mapped to corresponding mixed strategy using Boltzmann distribution policy. Hence the perception vector $(P^{(i)}(t))$ of network i is mapped to its corresponding mixed strategy, $p_i(t) = (p_1^{(i)}(t), p_2^{(i)}(t), ..., p_M^{(i)}(t))$ according to equation 2.

$$p_j^{(i)}(t) = \frac{e^{\gamma P_j^{(i)}(t)}}{\sum_{i=1}^{M} e^{\gamma P_j^{(i)}(t)}}, \forall j \in M \tag{2}$$

where γ is the temperature parameter in Boltzmann's distribution which controls the exploration of strategy space of a player. The value of γ changes over game stages as the experience of player increases. Initially the γ can be set to low value which emphasize each actions to be chosen with equal probability. Later the value of γ can be increased so that the stochastic exploration will be reduced and settle down in exploiting the bands that have high perception value.

After mapping the perception vector to corresponding mixed strategy, a stochastic action is taken for each network i, to decide the operating spectrum band for the next game stage. The players observe the payoff for the previously taken action and update their perception vector for current period. The updated perception for network i about a band j for stage $t+1$ is given in equation (3). According to the expression (3), the network i has already played t stages and recorded the perception vector over the stages. At the end of each stage t, the networks update their perception vector which will be used as decision criteria for choosing a band in the next game stage $(t+1)$. If network i has successfully acquired a band j by taking action $a_i(t) = j$ in game stage t, then the perception $(P_j^{(i)}(t))$ about the band j must increase in proportion to utility reward from that band for stage $(t+1)$. And for unsuccessful possession of band j will lead to decrease in perception value of that band which is expressed in the first case of eqn. (3). The perceptions of un-accessed bands in stage t remain unaltered. Algorithm 1 summarizes the distributed procedure for self-coexistence in homogeneous bands scenario using perception based learning model.

$$P_j^{(i)}(t+1) = \begin{cases} (1-\mu_t)P_j^{(i)}(t) + \mu_t U_{i,j}(t) & \textbf{if } a_i(t) = j \\ P_j^{(i)}(t) & \textbf{otherwise} \end{cases} \tag{3}$$

where $\mu_t \in (0,1)$ is the smoothing variable factor which changes over the stages. Initially the value of μ_t is set to be high by the system which allows the network players to explore the strategy space. Gradually, the value of μ_t is decreased so that the networks will settle down on the particular band whose perception value is high.

Algorithm 1: Perception Based Learning

```
Initialize γ and P_j^(i)(0) = 1/M for all networks i ∈ N and j ∈ M;
while stage t ≤ MaxT do
    for all network i ∈ N do
        Select a band j ∈ M based on its mixed strategy equation (3);
        Observe the utility reward for the stage t, U_{i,a_i(t)}(t);
        Update the perception (P_j^(i)(t+1)) for all bands j ∈ M according to equation (4);
        t ← t + 1;
    end
end
```

III. Self-Coexistence in Heterogeneous Bands

A. Game Settings for heterogeneous band based self-coexistence problem

When the spectrum chunks have different characteristics, the utilities returned from each band is distinct. Unlike homogeneous bands, here the utility rewards from each spectrum is assumed to be different. Assuming M contention-free bands that deliver utilities $u_1, u_2, ..., u_M$, to the networks on each exclusive access, the utility function for network i can be defined as:

$$U_i(a_i, a_{-i}) = \begin{cases} u_j & \text{if } a \neq a_i,\ \forall a \in a_{-i} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Though this problem seems to be fairly similar to previously discussed one, it has more importance from the rationality perspective of network players in a distributed uncoordinated wireless environment. As all network players are greedy and always look for self-betterment, they will try to acquire the band that gives highest reward. Similar thought among other player will lead them to go for the most valuable band and collide which results failed transmission by incurring unnecessary switching cost which eventually reduces the total payoff. Hence the networks need to adopt some strategy so that they will get a contention-free band of fair utility within few game stages and the system utility is maximized. It can only be achieved when networks will explore all available spectrum bands and use some experience based heuristic to learn about the availability of bands and its throughput. After sufficient exploration, the networks will neglect the bands that return low utility reward and exploit a band with fairly high utility reward that will maximize its cumulative utility. We propose a heuristic that uses the previously described perception learning model along with a regret minimization technique to conduct strategic analysis about spectrum choice.

B. Regret minimization model

The most valuable band is the center of attraction for all players, however no single network will be able to successfully grab the channel due to simultaneous access leads to collision and high penalty. Hence each network must play strategically by taking right action at right moment to reach the optimal convergence point where the payoff over long run will be maximized. Before achieving optimality, networks may get contention-free band that does not necessarily maximize the system utility, we call it as sub-optimal convergence point.

The following regret minimization [8] heuristic is proposed to achieve optimal system utility by N networks by acquiring

best N bands out of M-heterogeneous bands, assuming all networks know about the total available bands and the utility reward from each of the bands. To achieve this, the networks use regret matching technique to maintain the regret difference of actions that would have given more utility reward than currently taken action. Thus the strategy of choosing an action $\bar{\alpha}$ by network i in stage $(t+1)$, must be a function of average regret accumulated for the current action. This regret function will be able to lead the networks to choose actions that result high reward. In the same way, all other networks will choose for the highest utility band to operate, but due to collision no one will be able to operate on that band. To control these collisions, we must use the perception vector which will reduce the probability of choosing the same action over number of collisions. This will finally lead all networks to stick to the band that returns fairly high utility reward after certain stages. Thus strategy of choosing an action $\bar{\alpha}$ for stage $(t+1)$ should be a function of the regret difference $(R_i^{\bar{\alpha}}(t))$ and perception $(P_{\bar{\alpha}}^{(i)}(t))$ of network i upto game stage t which is presented in equation 5. The average regret $(R_i^{\bar{\alpha}})$ accumulated for network i, for all actions, $\bar{\alpha} \in A_i$ up to stage t is given by

$$R_i^{\bar{\alpha}}(t) = \left(\frac{1}{t}\right) \sum_{t'=1}^{t} \left[U_{i,\bar{\alpha}}(t') - U_{i,\alpha}(t')\right]$$

where $U_{i,\alpha}(t)$ is the utility reward to network i by choosing action $\alpha \in A_i$ at stage t.

The action for network i for stage $t+1$ can be taken based on the following probability distribution $p_i^{\bar{\alpha}}(t+1)$ which rely on the accumulated regret difference and perception about all actions over previous t stages. In eq. 5, the normalized regret difference contributes in leading the networks to choose higher utility bands, however the normalized perception value will control the number of collisions by reducing the probability of action $\bar{\alpha}$.

$$p_i^{\bar{\alpha}}(t+1) = \frac{R_i^{\bar{\alpha},+}(t)}{\sum_{\bar{\alpha}\in A_i}\left[R_i^{\bar{\alpha},+}(t)\right]} * \frac{P_{\bar{\alpha}}^{(i)}(t)}{\sum_{\bar{\alpha}\in A_i}\left[P_{\bar{\alpha}}^{(i)}(t)\right]} \quad (5)$$

where $R_i^{\bar{\alpha},+}(t) = max(R_i^{\bar{\alpha}}(t), 0)$

IV. Simulation and Results

In this section, the simulation results for the problem of self-coexistence in homogeneous and heterogeneous spectrum are reported. The simulations are carried out using Matlab version 7.9 with the following parameter settings. The total number of networks (N) in the game and spectrum bands are assumed to be 150. The utility reward (α) for homogeneous bands is assumed to be 1. To achieve a good convergence we allowed all networks to play for 300 stages at max. The exploration parameter (γ) is varied from 10 to 0.1. Initially γ is set as high value to explore the spectra for gathering knowledge about interference in the bands and gradually decreased so that networks will exploit the best perceived band.

To analyze the effect of switching cost (c) on overall system utility, we simulated for different values of switching cost and plotted the comparison in Fig. 1. In the competition, collision between some networks led them switching many times to find a free spectrum. Therefore, the switching cost will reduce the average payoff and perception vector of each action taken in the game stages. Thus more number of stages are required to find free spectrum band with high perception value. To show the variability of system utility with varying number of networks (N), we experimented by fixing the number of

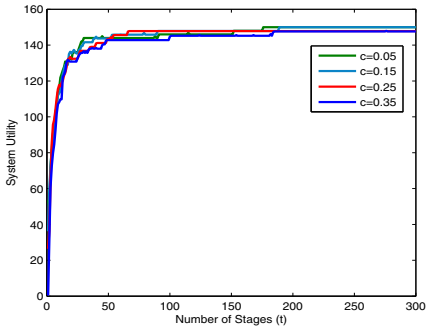

Fig. 1: Trade-off between System utility Vs switching cost(c)

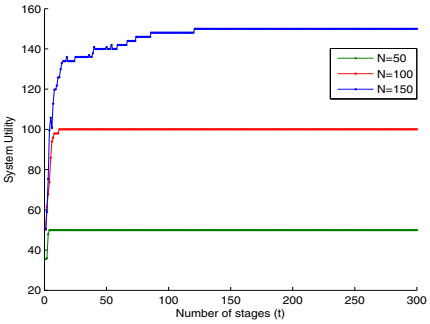

Fig. 2: System utility Vs Number of stages for varying N

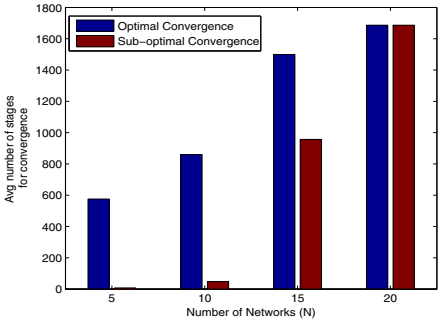

Fig. 3: Average number of stages for convergence Vs N

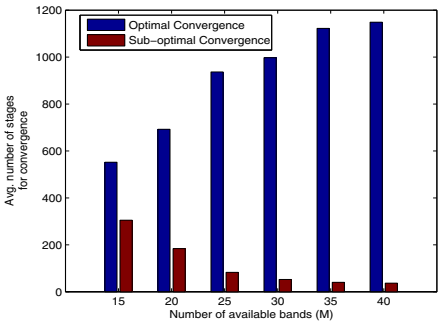

Fig. 4: Average number of stages for convergence Vs M

available spectrum bands (M) to 150, and utility reward of the spectrum bands to unity. For variable number of network players, we run the algorithm for 1000 times and the average result is reported in Fig. 2. We found that as the when number networks (N) is small compared to available spectrum bands (M), the networks can easily get a contention-free band within few stages of interaction because number of networks compete for many available resources is small. But when N is close to M, some networks switch indefinitely to acquire a contention-free band, therefore the convergence to highest system utility takes more number of game stages.

For self-coexistence in heterogeneous bands, we simulated the regret minimization based heuristic to achieve optimal system utility where all networks aim to occupy a band with fairly high utility reward. All the simulations are executed for 100 times, and the average results are presented.

To measure convergence in varying number of networks where 20 bands (M) are available, we experimented for various N values starting from 5. From Fig. 3, it can be observed that when the number of networks (N) in the game is fewer than available bands (M), finding a free spectrum band is relatively easy. To attain sub-optimal convergence, only few number of stages is required, however to achieve optimal system utility, the networks must compete hard among each other in a fairly large strategy space for enough number of stages to build better perception on a higher utility bands. But the important achievement here is that, the networks will ultimately acquire a band such that its gross utility is maximized. As the value of N approaches M, the number of game stages for converging to optimal system utility must rise to resolve the competition among each other and adapt to a band of preferably high utility reward.

To notice the effect of varying spectrum availability on number game stage required to converge, we fixed the number of networks (N) to 10 and varied M from 15 to 40. The reported result Fig. 4 conveys that the probability of finding a free band is more with increasing number of spectrum bands, because when the value of M surpasses the N, the CR networks find more free resources to use without contending others. Thus sub-optimal convergence period decreases with increase in M, however networks have to scavenge more to build their belief about the resources, thereby increases the foraging period to reach optimality. Finally we analyzed the number of stages for convergence with the following ratio mix: ratio of number of networks (N) and number of spectrum bands (M) is 0.5 and 0.75. From the Fig. 5, it can be observed that the number of stages required for optimal convergence is less with 50% ratio mix due to less intense competition than 75% ratio mix. And the similar rule applies to find sub-optimal convergence by CR networks where they need to compete more for exclusive spectrum possession.

V. Conclusions and Future Work

In this work, we have studied the issues of self-coexistence in cognitive radio network, which is an important aspect for maximizing spectrum utilization. We modeled the problem of

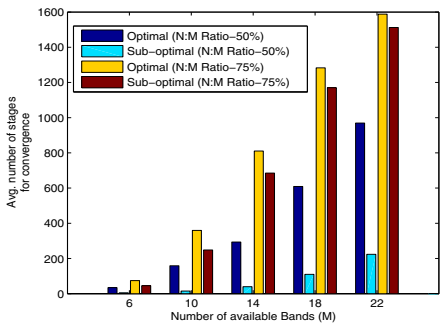

Fig. 5: Average number of stages for convergence Vs Number of Bands for 50% and 75% N:M ratio

self-coexistence as a standard multi-stage interaction game between N networks, competing for M homogeneous or heterogeneous bands, which was motivated from the famous optimal foraging model. We presented a perception based learning model which uses the past belief and utility reward to take decision to select bands in future stages. As shown in simulation results, this learning model helps networks to quickly learn and stick to best possible band according to the perception vector about all bands. We also define the importance of self-coexistence in heterogeneous bands. To achieve optimal system utility, a regret minimization heuristic was proposed that applies regret matching as well as perception model to maximize system utility. In future, we aim to apply exploration techniques to the regret minimization heuristic which will enable the networks to achieve maximum system utility in a variable spectrum opportunity environment and able to converge with minimum possible scavenging period.

References

[1] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty. Next generation/dynamic spectrum access/ cognitive radio wireless networks: a survey. Computer Networks, pp. 2127–2159, 2006.

[2] ETRI, Channel Management in IEEE 802.22 WRAN Systems, 2010.

[3] S. Sengupta, S. Brahma, M. Chatterjee, and S. Shankar, Enhancements to Cognitive Radio Based IEEE 802.22 Air-Interface. IEEE ICC, pp. 5155–5160, 2007.

[4] S. Sengupta, R. Chandramouli, S. Brahma, and M. Chatterjee, A game theoretic framework for distributed self-coexistence among IEEE 802.22 networks. In proceedings of IEEE Global Communications Conference (GLOBECOM), pp. 1–6, Nov.-Dec. 2008.

[5] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. Econometrica, pp. 1127–1150, 2000.

[6] K. Bian and J. Park, A Coexistence-Aware Spectrum Sharing Protocol for 802.22 WRANs. In Proceedings of International Conference on Computer Communications and Networks (ICCCN '09), pp. 1–6, 2009.

[7] M. Yoo and S. Hwang, A Self-Coexistence Method for the IEEE 802.22 Cognitive WRAN. In Recent Researches in Automatic Control, Systems Science and Communications.

[8] J. R. Marden, G. Arslan, and J. S. Shamma, Regret based dynamics: convergence in weakly acyclic games. In Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems (AAMAS '07), pp. 42:1–8, 2007.

[9] C.J.Watkins, Models of Delayed Reinforcement Learning. PhD thesis, Psychology Department, Cambridge University, 1989.

[10] A. Bab, R. Brafman, Multi-agent reinforcement learning in common interest and fixed sum stochastic games: an experimental study. Journal of Machine Learning Research, 9: 2635-2675, 2008.

[11] C. L. Gass, J. S. E. Garrison, Energy regulation by traplining hummingbirds. Functional Ecology, 13: 483492, 1999.