This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2008 workshop proceedings.

Cooperation and Learning in Multiuser Opportunistic Spectrum Access

Hua Liu§, Bhaskar Krishnamachari§, Qing Zhao‡
hual@usc.edu, bkrishna@usc.edu, qzhao@ece.ucdavis.edu
§Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089
‡Department of Electrical and Computer Engineering, University of California, Davis, CA 95616

***Abstract*— We consider how two secondary users should interact to maximize their total throughput in a two-channel sensing-based opportunistic spectrum access network where spectrum opportunities are time varying and spatially inhomogeneous. By modeling the occupancy of the primary users as discrete-time Markov chains, we obtain the optimal dynamic coordination policy using a partially observable Markov decision process (POMDP) solver. We also develop several tractable approaches - a cooperative multiuser approach based on explicit communication between the secondary users, a learning-based approach involving use of collision feedback information, and a single-user approach based on uncooperative independent decisions. As a baseline we consider the static partitioning policy where both users are allocated a single channel of their own. Simulations comparing the performance of these strategies yield several interesting findings: that significant improvements over static partitioning are possible with the optimal scheme; that the cooperative multiuser approach shows near-optimal performance in all cases; that there are scenarios when learning through collision feedback can be beneficial; and that the single-user approach generally shows poor performance.**

I. Introduction

A promising approach for the efficient utilization of the radio frequency spectrum is opportunistic spectrum access (OSA), where secondary users sense for the presence of primary users and transmit only when suitable opportunities arise.

A basic component of OSA is a sensing strategy at the MAC layer for spectrum opportunity tracking. Since a secondary user may not be able to sense all channels in the spectrum simultaneously, a sensing strategy for intelligent channel selection is crucial to track the rapidly varying spectrum opportunities.

By modeling primary users' channel occupancy as a Markov process, the design of sensing strategies is formulated as a partially observable Markov decision process (POMDP) in [1], [2], where the objective is to maximize the throughput of an individual selfish secondary user. The interaction among secondary users is not taken into account in the design of the sensing strategy.

The optimal sensing strategy designed for individual users is, however, suboptimal in terms of network throughput. Our goal in this study is to investigate whether and how contending secondary users should cooperate and learn from "mistakes" (collisions) in order to maximize the network throughput.

Intuitively, to maximize the network throughput, secondary users should seek spectrum opportunities in different channels, which avoids collision among secondary users and exploits fully the spectrum opportunities offered by multiple channels. This suggests a channel partition strategy which preassigns each secondary user a distinct set of channels for use (assuming the number of channels is larger than the number of secondary users). If all secondary users are fully backlogged and affected by the same set of primary users (thus experiencing the same spectrum opportunities), this channel partition strategy yields an efficient solution.

In this paper, we show that when secondary users are affected by different sets of primary users, it is no longer optimal to partition the channels among secondary users and more sophisticated strategies are called for. For analytical tractability and in order to obtain general insights, we consider a simple setting with two interfering secondary users, each affected and observing different primary users. We develop and compare a number of different strategies that the two secondary users can adopt to make decisions at each step on which channel they should sense. The sensing decision not only helps obtain immediate rewards in terms of free channels, but also provides fresh information regarding the primary users' occupancy which in turn can help improve future decisions. If a secondary user senses a channel to be free at any time slot, it transmits during that slot. Depending on the strategy adopted, it can happen that both users sense the same channel to be free and transmit simultaneously, which is assumed to result in a collision. In a multi-user setting, even such collisions can carry useful information for the secondary users to learn about each other's behavior.

We formulate the problem as a finite-horizon POMDP, in which the history of previous observations is summarized by a set of beliefs regarding the presence of primary users on the channels that both secondary users maintain and update. Obtaining a solution to a POMDP is in general intractable, and even with two users and two channels, solutions may be difficult to obtain for large horizons. To address this, we develop and evaluate three variants of myopic policies, in which the users act to maximize their immediate expected reward, given their beliefs. These variants are distinguished by the information available to the two secondary users regarding each others' beliefs. The first, which we term "cooperative multi-user," assumes that users exchange their beliefs at each time slot. In the second, which we term "learning-based multiuser," each user maintains an estimate of the other user's belief implicitly learned and updated based on occurrence of collisions. In the third scheme, which we term "single-user," each user acts uncooperatively and entirely ignores all observations pertaining to the other user.

We evaluate the performance of these schemes through a comprehensive set of simulations. For a baseline, we compare the above schemes with the simple strategy of statically partitioning the two channels between the two users. The results we obtain shed light on the role of cooperation and learning in multiuser OSA under different scenarios.

Related Work Existing work within the POMDP framework for OSA includes [1], [2], [3], all focusing on the single-user setting. In [4], the POMDP framework is extended to a multiuser setting for spatially homogeneous spectrum opportunities. Random traffic arrivals at each secondary user is considered in [4], which renders a fixed channel partition among users inefficient, and a randomized dynamic sensing policy is proposed to address the tradeoff between choosing the channel that is most likely to offer a spectrum opportunity and avoiding other competing secondary users.

Under the assumption of time-invariant (or slowly varying) spectrum opportunities, spectrum sharing among competing secondary users has been addressed using graph coloring theory [5], [6] and game theory [7], [8], [9], [10]. These approaches assume perfect knowledge of spectrum opportunities at any time and location over all the channels. In [11], only statistical information about spectrum opportunities (the probability that each channel presents an opportunity to a secondary user) is assumed. The optimal sensing actions for single-stage interactions among secondary users are obtained for both spatially homogeneous and inhomogeneous spectrum opportunities.

Another related work is [12], which considers multiple secondary users sharing spatially homogeneous spectrum opportunities. The time variation of spectrum opportunities is also modeled by discrete-time Markov chains. Bidding policies for secondary users are developed within a stochastic game framework by assuming a central coordinator who has the knowledge of the occupancy state of all the channels.

II. Problem Definition

A. Network Model

Assume the spectrum is divided into M independent channels that are leased to a time synchronized slot-based primary network with multiple primary users. In the secondary network, there are N users; each chooses one channel to sense at the beginning of each time slot and transmits if an opportunity exists (i.e. if there is no primary user occupying the channel in that slot). In general, depending on the geographical locations of the nodes, each secondary user i can be in the range of a different set of primary users.

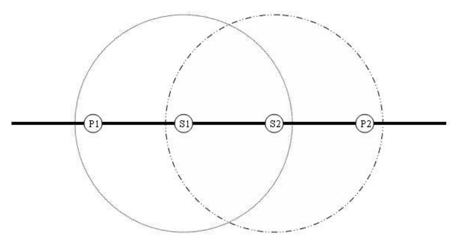

Fig. 1. Illustration of Scenario

Figure 1 shows the scenario considered in this work - two secondary users ($N = 2$) contending with each other but perceiving different primary users. Spectrum opportunities for secondary users S1 and S2 are determined by two independent sets of primary users (P1 and P2, respectively). We consider $M = 2$ channels in this study. The availability of channel j for secondary user i is modeled as a two-state discrete time Markov chain with state $s_j^i(t)$ indicating whether an opportunity ($s_j^i(t) = 1$) exists for secondary user i in channel j at slot t. To simplify notation and without loss of generality, we assume that for each secondary user i, the states of the two channels form identical Markov chains with one-step transition probability denoted by $p_{j,k}^i$ where $j, k \in \{0, 1\}$.

B. POMDP Formulation

We formulate the opportunistic channel access problem as a POMDP represented by the tuple (S, A, O, P^s, P^o, R) given below.

- S is the combination of channel availability states for every secondary user. $S = \times_{i\in N, j\in M} s^i_j$, i.e. $S = \{0,1\}^{MN}$.
- A is the joint action profile for all the users. The set of actions in this problem are the decision for each user on which channel to sense at the beginning of each time slot. $A = \{1,2\}^N$ for our case where $M = 2$. We use $a^i(t)$ to denote the action for secondary user i in time slot t.
- O is the observation space. We denote by $O = \times_{i\in N} O^i$ the joint observations of all secondary user i and $o(t) = (o^1(t), ..., o^n(t))$ is a joint observation in time slot t. For each user i, there exist three kinds of observations in a time slot t: busy, collision, and success. Busy means that the channel was occupied by the primary user in this slot so that the secondary user must defer to the next slot. If the channel is sensed to be free, the secondary user transmits a packet. If two secondary users transmit simultaneously it results in a collision, else the transmission results in a success.
- P^s is a set of Markovian state transition probabilities. $P^s(s, s') = Pr\{s(t+1) = s'|s(t) = s\}$ denotes the probability of being at state s' at time slot $t+1$ when given that at time slot t, spectrum state is at state s. Note that the spectrum state transition of primary users is independent of the actions made by secondary users.
- P^o represents the probability that action a for state s at time slot t will give observation o, i.e. $P^o(s, a, o) = Pr\{o(t) = o|s(t) = s, a(t) = a\}$.
- R represents the reward function mapping from the observation space O to real numbers; $R^i(t)$, the reward for secondary user i in time slot t defined as follows:

$$R^i(t) = \begin{cases} 0 & \text{if } o^i(t) \text{ is busy or collision} \\ 1 & \text{if } o^i(t) \text{ is success} \end{cases}$$

The system reward is the expected reward summation for all the secondary users.

$$R = E(\sum_{t=1}^{T}\sum_{i=1}^{N} R^i(t))$$

where T is the total time slot, also named horizon.

A sensing policy π is a policy to decide for each secondary user what action to take in each time slot. R_π denotes the system reward of a policy π, which is defined as the expected reward summation of horizon T for all users. In evaluating different policies, we will use the normalized reward $r_\pi = \frac{R_\pi}{T}$ which represents average per-slot throughput as the evaluation criterion.

C. Decision Cycle

In each time slot, each secondary user experiences the following three phases.

1) Decision Phase: at the beginning of slot t, secondary user i takes the action $a^i(t)$, deciding which channel to sense in this time slot.
2) Transmission Phase: upon taking the sensing action, the user transmits if an opportunity exists. The observation $o^i(t)$ follows, determined jointly by the primary-users' occupancy in time slot t and the actions of both users $a^1(t), a^2(t)$.
3) Reward Phase: If transmission is successful in that slot a unit reward is accrued, determining $R^i(t)$.

D. Belief Vectors and Myopic Policy

Due to limited sensing, each secondary user cannot directly observe the exact spectrum status of both channels at each time slot. Instead it obtains a distribution (belief) on the non-observed channel from the history of sensing results, taking into account the Markov process governing the evolution of the channel state.

Definition 1: A belief vector for user i at time slot t $\Upsilon^i(t)$ is a M-dimension vector $(\upsilon^i_1(t), \upsilon^i_2(t), ..., \upsilon^i_M(t))$ for in a M-channel system. Element $\upsilon^i_j(t)$ is channel i's available probability for the secondary user j in time slot t.

Belief vectors are updated at the end of each time slot. Specifically, for cognitive radio opportunistic channel accessing problem with two channels in the system, the belief vector updating rules are represented in Figure 2. Please note that the two secondary users have different belief vectors due to their different observation histories. They do not know the other user's belief unless through explicit cooperation or learning, as discussed in Sec. III. We, however, assume that a secondary user and its intended receiver can maintain the same belief vector so that they will choose the same channel for communication. This can be achieved by ensuring that the transmitter and the receiver use common observations in the belief update, where common observations are obtained when the transmitter and its receiver are affected by the same set of primary users or through feedbacks such as acknowledgement from the receiver to the transmitter (see [1]).

Fig. 2. Belief Vector Updating Rules

$$\upsilon_j^i(t+1) = \begin{cases} p_{1,1}^i & \text{if } a^i(t) = j \text{ and } s_j^i = 1 \\ p_{0,1}^i & \text{if } a^i(t) = j \text{ and } s_j^i = 0 \\ p_{1,1}^i \upsilon_j^i(t) + p_{0,1}^i(1 - \upsilon_j^i(t)) & \text{if } a^i(t) \neq j \end{cases}$$

III. Tractable Policies

In this section, we develop and compare three approaches to multiuser OSA that involve different degrees of cooperation among the secondary users. While the optimal policy in each approach can be obtained, the complexity grows exponentially with the horizon length. We thus focusing on myopic solutions for all approaches; the main objective is to explore the role of cooperation and learning through performance comparison of these approaches. A static channel partition strategy is also presented as the baseline performance.

A. Static Partition Strategy

The static partition strategy is a straightforward approach to completely avoid collision among secondary users: to assign a fixed channel to each secondary user. Note that we assume the two channels are identical, therefore it doesn't matter which channel is assigned to which user. Without loss of generality, we assign channel i to user i:

$$a^i(t) = i \qquad \text{for } i = 1, 2$$

B. The Single-User Approach

In single user myopic approach, each secondary user executes an independent myopic policy, ignoring the presence of the other user.

$$a^i(t) = arg \max_{a=1,\dots,N} E(R^i(t)|\Upsilon^i(t))$$

Obtaining the myopic action under this approach is trivial - each secondary user independently picks the channel for which it has a higher belief of seeing an opportunity.

The major drawback of the single user approach is that it does not take collision among secondary users into consideration.

C. Cooperative Multi-User Approach

This cooperative policy allows secondary users to exchange their belief vectors at each time slot and use these information to generate consistent actions:

$$a^i(t) = arg \max_{a=1,\dots,M} E(\sum_i R^i(t)| \bigcup_{j\in N} \Upsilon^j(t))$$

Computing the optimal action for the cooperative multi-user myopic approach is somewhat involved.

Let α_i denote the probability for secondary user i to sense channel 1. Since we assume that in each time slot, a secondary user must choose a channel to sense, the probability for secondary user i to sense channel 2 is $1 - \alpha_i$.

The objective of maximizing the expected immediate system reward in time slot t can be formulated as the following:

$$\begin{aligned} \max \ & \alpha_1(1-\alpha_2)(\upsilon_1^1(t) + \upsilon_2^2(t)) \\ & +(1-\alpha_1)\alpha_2(\upsilon_2^1(t) + \upsilon_1^2(t)) \\ & +\alpha_1\alpha_2((1-\upsilon_1^1(t))\upsilon_1^2(t) + (1-\upsilon_1^2(t))\upsilon_1^1(t)) \\ & +(1-\alpha_1)(1-\alpha_2)((1-\upsilon_2^1(t))\upsilon_2^2(t) \\ & +(1-\upsilon_2^2(t))\upsilon_2^1(t)) \\ such\ that \quad & 0 \leq \alpha_i \leq 1 \qquad for\ i = 1, 2 \end{aligned}$$

It can be shown that the solution for this maximization occurs at an extreme point (i.e. when each user picks one of the two channels with probability one).

Table 1 gives the solution of α_1,α_2 value and corresponding expected rewards in one time slot. Using its own current belief vector and that of the other secondary user, each user can compute all four possibilities and pick the action that corresponds to the maximum expected reward (the other user will also do the same as it has consistent information).

TABLE I

Table for Maximized Expected Rewards

α_1	α_2	Expected Total Rewards
0	0	$(1-\upsilon_2^1(t))\upsilon_2^2(t) + (1-\upsilon_2^2(t))\upsilon_2^1(t)$
0	1	$\upsilon_2^1(t) + \upsilon_1^2(t)$
1	0	$\upsilon_1^1(t) + \upsilon_2^2(t)$
1	1	$(1-\upsilon_1^1(t))\upsilon_1^2(t) + (1-\upsilon_1^2(t))\upsilon_1^1(t)$

D. Learning-based Multi-User Approach

Collision among secondary users reduces the system performance. However, collision information can also provide accurate belief vector estimation for secondary users. In this approach, the secondary user not only maintains a belief vector of itself according to updating rules described in section II-D, but also maintains an estimated belief vector for the other secondary user by learning from collision events as per rules depicted in Figure 3. Note that no message exchange between the two users is needed in this approach.

In the description of these rules, $collision(i, -i)$ is a Boolean expression to indicate whether there is a collision between two secondary users in time slot t, and $\hat{a}^{-i}(t)$ is the inferred action of user $-i$ by user i using estimated belief vector $\hat{v}^{-i}(t)$ in time slot t.

There are four cases in estimating the updated belief vector: a) collision detected between secondary users; b) user i successfully transmits in this time slot without collision with the other user $-i$, while the other user $-i$ should choose same channel to transmit according to previous estimated belief vectors; c) and d) both represent the case when user i and user j choose different channels to transmit in this time slot.

The actions are obtained in this approach using the same method as cooperative multi-user myopic policy except that the estimated belief vector is used for that other user (instead of the exact belief vector).

IV. Simulation Results

Table II shows the results obtained from a comprehensive set of simulations in MATLAB performed to evaluate and compare the various approaches. Since we have fixed the scenario to involve exactly two users and two channels, the key parameters that determine relative performance are the transition matrices characterizing the two users. These are summarized by four numbers, namely, $p^1_{0,1}$, $p^1_{1,0}$, $p^2_{0,1}$, and $p^2_{1,0}$. We considered low (0.15) and high (0.95) values for these parameters and ran simulations over 10 cases that capture all possible unique settings of these values (barring symmetric settings). In these simulations the belief vectors for each user are initialized to the corresponding stationary distributions. In all cases the horizon length is 1000, except in cases 4 and 10, where the optimal solution could only be found for a short horizon length of 8 and 6, respectively. The results presented are all averaged over 100 runs.

We have also obtained the optimal policy with global information using a POMDP solver [13] to provide a performance benchmark. The per-slot network throughput for different approaches are given in Table II.

Table II yields the following interesting insights:

- The cooperative approach, which uses a tractable myopic policy, is almost indistinguishable from optimal in nearly all cases (except case 1 where it is only about 5 % off).
- The single-user approach generally performs worse (sometimes drastically, as in case 5) than even the baseline of static partition, suggesting that some degree of cooperation is essential.
- While the learning-based approach (which involves no message exchange) does not always perform as well as the cooperative scheme, there are cases (1, 3, 9) where it is better than fixed partition.

Based on these results we hypothesize that the conditions under which the learning-based approach can outperform static partition are when the following conditions hold: (1) one of the two secondary users is dominant in the sense that it has a higher stationary probability of finding an opportunity compared to the other user (i.e., user 1 is dominant if $\frac{p^1_{0,1}}{p^1_{0,1}+p^1_{1,0}} > \frac{p^2_{0,1}}{p^2_{0,1}+p^2_{1,0}}$), and (2) this dominant user prefers to switch channels in order to maximize its own throughput (this would occur for user i if $p^i_{1,0}$ is large). Intuitively, under these settings, the static partition strategy has room for improvement since the gains that the dominant user can make by switching channels when needed cannot be offset by the long-term average reward obtained by the non-dominant user staying on a single channel.

To verify this hypothesis, we ran some additional detailed simulations covering a wider range of cases, shown in figure 4. The plot has 100 points obtained by varying $p^1_{0,1}$ and $p^2_{0,1}$ from 0.1 to 1 with a step-size of 0.1, $p^1_{1,0} = p^2_{1,0} = 0.9$. In the plot a blue cross point means that learning-based approach outperformed partitioning; red circles mean partitioning performed better. A green star implies that they have the same performance. The points are obtained by comparing the average of 100 runs over a horizon length of 300.

Condition (2) of our hypothesis holds in this plot because of the high values of $p^i_{1,0}, p^2_{1,0}$, and condition (1) holds in regions away from the 45-degree diagonal. We do indeed see that the learning approach performs better than partition when both conditions hold, supporting the hypothesis.

V. Conclusion

The results from this study provide some useful general lessons regarding multiuser sensing for opportunistic sensing. They indicate that some degree of cooperation is essential - secondary users cannot afford to completely ignore the presence of other secondary users and use a single-user approach. The cooperative myopic approach shows near-optimal performance in our simulations and is generally substantially better than static partitioning, but it requires explicit belief exchange among users at each step. Thus this strategy may be most useful for

Fig. 3. Rule for Estimating the Other User's Belief Vector in the Learning-based Approach

$$\hat{v}_j^{-i}(t+1) = \begin{cases} p_{1,1}^{-i} & \text{if } a^i(t) = j \text{ and } collision(i,-i,t) = True \\ p_{0,1}^{-i} & \text{if } a^i(t) = j \text{ and } \hat{a}^{-i}(t) = j \text{ and } s_j^i = 1 \text{ and } collision(i,-i,t) = False \\ p_{1,1}^{-i}\hat{v}_j^{-i}(t) + (1-\hat{v}_j^{-i}(t))p_{0,1}^{-i} & \text{if } a^i(t) = j \text{ and } \hat{a}^{-i}(t) \neq j \text{ and } s_j^i = 1 \text{ and } collision(i,-i,t) = False \\ p_{1,1}^{-i}\hat{v}_j^{-i}(t) + p_{0,1}^{-i}(1-\hat{v}_j^{-i}(t)) & \text{if } a^i(t) \neq j \end{cases}$$

TABLE II

NETWORK THROUGHPUT COMPARISON

Case	$[p_{0,1}^1, p_{1,0}^1, p_{0,1}^2, p_{1,0}^2]$	Optimal	Cooperative	Learning-based	Partition	Single User
1	[0.95, 0.95, 0.95, 0.95]	1.32	1.28	1.04	1.00	0.92
2	[0.95, 0.95, 0.95, 0.15]	1.59	1.59	1.14	1.36	0.98
3	[0.95, 0.95, 0.15, 0.95]	0.86	0.86	0.76	0.63	0.76
4	[0.95, 0.95, 0.15, 0.15]	1.26*	1.28	1.00	0.99	0.91
5	[0.95, 0.15, 0.95, 0.15]	1.74	1.74	1.41	1.72	1.01
6	[0.95, 0.15, 0.15, 0.95]	1.00	1.00	0.91	0.99	0.89
7	[0.95, 0.15, 0.15, 0.15]	1.54	1.54	1.33	1.36	0.96
8	[0.15, 0.95, 0.15, 0.95]	0.29	0.29	0.27	0.27	0.27
9	[0.15, 0.95, 0.15, 0.15]	0.80	0.80	0.72	0.63	0.71
10	[0.15, 0.15, 0.15, 0.15]	1.33*	1.18	0.91	1.00	0.90

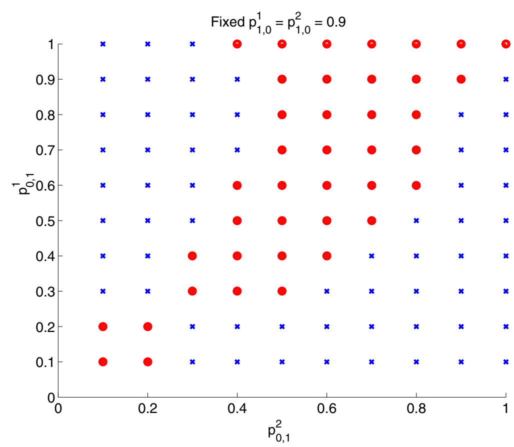

Fig. 4. A detailed performance comparison of the learning-based multi-user approach versus the static partitioning approach

settings such as high-bandwidth applications where the control overhead of communicating beliefs is small compared to the data size. We have also identified conditions where implicit learning based on collision feedback may be effective compared to static partitioning.

REFERENCES

[1] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework," *IEEE Journal on Selected Areas in Communications: Special Issue on Adaptive, Spectrum Agile and Cognitive Wireles Networks*, April 2007 (also see *Proc. of DySPAN*, pp. 224 - 232, Nov. 2005).

[2] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *submitted to IEEE Transactions on Information Theory*, Feb. 2007; available at http://www.ece.ucdavis.edu/∼qzhao/Journal.html (also see *Proc. of IEEE Asilomar Conference on Signals, Systems, and Computers*, Oct. 2006).

[3] Q. Zhao and B. Krishnamachari, "Structure and optimality of myopic sensing for opportunistic spectrum access," in *Proc. of IEEE Workshop on Toward Cognition in Wireless Networks (CogNet)*, June, 2007.

[4] K. Liu, Q. Zhao, and Y. Chen, "Distributed Sensing and Access in Cognitive Radio Networks," in *Proc. of 10th International Symposium of Spread Spectrum Techniques and Applications* (Invited paper), August, 2008.

[5] H. Zheng and C. Peng, "Collaboration and Fairness in Opportunistic Spectrum Access," in *Proceedings of IEEE International Conference on Communications (ICC)*, 2005.

[6] W. Wang and X. Liu, "List-coloring based channel allocation for open-spectrum wireless networks," in *Proc. of IEEE VTC*, 2005.

[7] M. Halldorsson, J. Halpern, L. Li, and V. Mirrokni, "On spectrum sharing games," in *Proc. of the 23rd Annual ACM Symposium on Principles of Distributed Computing*, pp. 107–114, 2004.

[8] Z. Ji and K. J. R. Liu, "Dynamic spectrum sharing: A game theoretical overview," *IEEE Communications Magazine*, vol. 45, pp. 88 – 94, May 2007.

[9] Y. Xing, C. N. Mathur, M. A. Haleem, R. Chandramouli, and K. P. Subbalakshmi, "Dynamic spectrum access with QoS and interference temperature constraints," *IEEE Transactions on Mobile Computing*, vol. 6, pp. 423 – 433, Apr. 2007.

[10] J. E. Suris, L. A. DaSilva, Z. Han, and A. B. MacKenzie, "Cooperative game theory for distributed spectrum sharing," in *Proceedings of IEEE International Conference Communications (ICC)*, pp. 52825287, June 2007.

[11] H. Liu, and B. Krishnamachari, "Randomized Strategies for Multi-User Multi-Channel Opportunity Sensing," Workshop on Cognitive Radio Networks, IEEE CCNC, (Invited paper), January 2008.

[12] F. Fu and M. van der Schaar, "Learning for Dynamic Bidding in Cognitive Radio Resources," submitted.

[13] A.R. Cassandra, POMDP solver in C, Version 4.0, http://pomdp.org/pomdp/code/index.shtml