

Dynamic spectrum access with learning for cognitive radio

Jayakrishnan Unnikrishnan and Venugopal V. Veeravalli

Department of Electrical and Computer Engineering, and Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

Abstract—We study the problem of cooperative dynamic spectrum sensing and access in cognitive radio systems as a partially observed Markov decision process (POMDP). Assuming Markovian state-evolutions for the primary channels, we propose a greedy channel selection and access policy that satisfies an interference constraint and also outperforms some existing schemes in average throughput. When the distribution of the signal from the primary is unknown and belongs to a parameterized family, we develop an algorithm that can learn the parameter of the distribution still guaranteeing the interference constraint. This algorithm also outperforms the popular approach that assumes a worst-case value for the parameter thus illustrating the suboptimality of the popular worst-case approach.

I. INTRODUCTION

In order to maximize throughput, a cognitive radio system must have an efficient channel selection and access policy. The cognitive radio (also called the ‘secondary user’) needs to decide what is the best strategy for selecting licensed (or ‘primary’) channels for sensing and access. The sensing and access policies should jointly ensure that the probability of interfering with the primary’s transmission meets a given constraint. This paper deals with the design of such a joint sensing and access policy assuming a Markovian structure for the primary spectrum usage on the channels being monitored. In most of the existing schemes in the literature in this field, the authors either assume error-free observations of the channel states [1], [2], [3] or assume that the channel states are learned based on the ACK signals transmitted from the secondary’s receivers [4]. We adopt a different strategy in which we use the observations made in each slot to learn the occupancies of the different channels. We introduce this problem in section II and describe our proposed solution in section III.

In the second part of the paper, we propose and study a more practical problem which arises when the secondary users are not aware of the exact distributions of the signals that they receive from the primary transmitters. We model this uncertainty by assuming that these signals have distributions parameterized by unknown random parameters. We develop a scheme that learns these parameters online, still satisfying a constraint on the probability of interfering with the primary signals. Through simulations, we show that this scheme gives improved performance over the naive scheme that assumes worst-case values for the unknown parameters. We elaborate

on this problem in section IV. We present our simulation results along with some comparisons in section V and our conclusions in section VI.

II. PROBLEM STATEMENT

We consider a slotted system in which a group of secondary users located close to each other monitor a set \mathcal{C} of primary channels that have equal bandwidth B and are statistically identical and independent in terms of primary usage. The state of each channel a at any time slot k , represented by variable $S_a(k)$, could be either 0, meaning the channel is free for secondary access, or 1, meaning the channel is occupied by some primary user. The state evolution follows a stationary Markov chain. The secondary users can cooperatively sense any one channel in \mathcal{C} in each slot, and can access any one of the $L = |\mathcal{C}|$ channels in the same slot. When the secondary users access a channel that is free during a given time slot, they receive a reward proportional to the bandwidth of the channel that they access. The objective of the secondary users is to select the channels for sensing and access in each slot in such a way that their total expected reward accrued over all slots is maximized subject to a constraint ζ on interfering with potential primary transmissions every time they access a channel.¹ Since the secondary users do not have explicit knowledge of the states of the channels, the resultant problem is a constrained partially observable Markov decision process (POMDP) problem.

The Markovian statistics of the primary channels are assumed to be known by the secondary users. For the problem studied in sections II and III, we assume that the knowledge of these statistics is not accurate. The system could be one where the secondary users adaptively learn these usage statistics. Hence, the secondary users could make use of the knowledge of these statistics to improve their accrued reward, but should meet the constraint on the probability of interfering with primary transmissions without relying on these statistics.

The secondary system includes a decision center that has access to all the observations² made by the cooperating secondary users. The decisions about which channels to sense and access in each slot are made at this decision center. When channel a is sensed in slot k , we use $X_a(k)$ to denote the

¹We assume that the secondary users have some given fixed scheduling policy to decide which user accesses the primary channel every time they determine that a channel is free for access.

²Quantized versions are sufficient. However, we do not discuss this case to avoid details about discrete-valued observation variables.

vector of observations made by the different cooperating users on channel a in slot k . The statistics of these observations are assumed to be time-invariant. The observations on channel a have distinct joint probability density functions f_0 and f_1 under H_0 (i.e. $S_a(k) = 0$) and H_1 (i.e. $S_a(k) = 1$) respectively. The collection of all observations up to slot k is denoted by \underline{X}^k , and the collection of observations on channel a up to slot k is denoted by \underline{X}_a^k . We use u_k to denote the channel that is sensed in slot k , and K_a^k to represent the time slots up to slot k when channel a was sensed. The decision to access channel a in slot k is denoted by a binary variable $\delta_a(k)$, which takes value 1 when channel a is accessed in slot k , and 0 otherwise.

Whenever the secondary users access a free channel in some time slot, they get a reward B equal to the common bandwidth of each channel in \mathcal{C} . The interference constraint on each channel a in each slot k can be expressed as:

$$P(\{\delta_a(k) = 1\}|\{S_a(k) = 1\}) < \zeta$$

Since the knowledge about the primary users' usage statistics is not necessarily accurate, the secondary users should satisfy this constraint using only the information gained in the current slot. Hence it follows that in each slot k the secondary users should be allowed to access only the channel u_k they observe in slot k . Furthermore, if $u_k = a$, the decision to access must be based on a binary hypothesis test [5] between the two hypotheses H_0 and H_1 performed on the observation $\underline{X}_a(k)$. The optimal test [5] is to compare the joint log-likelihood ratio $\mathcal{L}(\underline{X}_a(k))$ given by:

$$\mathcal{L}(\underline{X}_a(k)) = \log \left(\frac{f_1(\underline{X}_a(k))}{f_0(\underline{X}_a(k))} \right)$$

to some threshold Δ chosen to satisfy:

$$P(\{\mathcal{L}(\underline{X}_a(k)) < \Delta\}|\{S_a(k) = 1\}) = \zeta \quad (1)$$

and the optimal access decision would be to access the sensed channel whenever the threshold exceeds the joint log-likelihood ratio. Hence

$$\delta_a(k) = \mathcal{I}_{\{\mathcal{L}(\underline{X}_a(k)) < \Delta\} \cap \{u_k = a\}} \quad (2)$$

and the reward obtained in slot k can be expressed as:

$$\hat{r}_k = B \mathcal{I}_{\{S_{u_k}(k)=0\} \cap \{\mathcal{L}(\underline{X}_{u_k}(k)) < \Delta\}} \quad (3)$$

where \mathcal{I}_E represents the indicator function of event E . Thus the problem of spectrum sensing and access reduces to choosing the optimal channel to sense in each slot. Whenever the secondary users sense some channel and make observations with log-likelihood ratio lower than the threshold, they are free to access that channel. Thus we have converted the constrained POMDP problem into an unconstrained POMDP problem. Since we do not know the exact number of slots over which we need to optimize the expected accrued reward, we introduce a discount factor $\alpha \in (0, 1)$ and aim to solve the infinite horizon dynamic program with discounted rewards. That is, we seek the sequence of channels $\{u_0, u_1, \dots\}$, such that $\sum_{k=0}^{\infty} \alpha^k E[\hat{r}_k]$ is maximized, where the expectation is performed over the

random observations and channel state realizations. This is equivalent to maximizing the total expected reward assuming that the number of slots over which the secondary users need to use the spectrum is distributed as a geometric random variable with mean $\frac{1}{1-\alpha}$. By employing iterated conditioning, it can be shown that we can redefine our problem such that the reward in slot k is now given by:

$$r_k = B(1 - \hat{\epsilon}) \mathcal{I}_{\{S_{u_k}(k)=0\}} \quad (4)$$

$$\text{where } \hat{\epsilon} = P(\{\mathcal{L}(\underline{X}_a(k)) > \Delta\}|\{S_a(k) = 0\}) \quad (5)$$

and the optimization problem is equivalent to maximizing $\sum_{k=0}^{\infty} \alpha^k E[r_k]$. In (4), B and $\hat{\epsilon}$ are constants that do not depend on the channel selected since all channels are assumed to be identical in bandwidth and primary signal characteristics.

III. DYNAMIC PROGRAMMING

The state of the system at time k is given by the vector of states of the channels in \mathcal{C} : $\underline{S}(k) = (S_1(k), S_2(k), \dots, S_L(k))^T$ which have independent Markovian evolutions. We use $I_k := (\underline{X}^k, u^k)$ to denote the net information about past observations and decisions available at the end of slot k . The channel, u_k to be sensed in slot k is decided in slot $k-1$ and is a deterministic function of I_{k-1} . The expected reward obtained in slot k is a function of $\underline{S}(k)$ and u_k as given by (4). We seek the sequence of channels $\{u_0, u_1, \dots\}$, such that $\sum_{k=0}^{\infty} \alpha^k E[r_k]$ is maximized.

Under this formulation it can be easily verified that this problem is essentially a standard dynamic programming problem with imperfect observations. It is known [6] that for such a POMDP problem, a sufficient statistic at the end of any time slot k , is the conditional probability distribution of the system state $\underline{S}(k)$ conditioned on all the past observations and decisions and is given by $P(\{\underline{S}(k) = \underline{s}\}|I_k)$. Since the channel evolutions in our problem are independent of each other, this conditional probability distribution is equivalently represented by the set of *beliefs* about the occupancy states of each channel, i.e., the conditional probability of occupancy of each channel at time k , conditioned on all the past observations on channel a and on the time slots when channel a was sensed. We use $p_a(k)$ to represent the belief about channel a at end of slot k , given by $p_a(k) = P(\{S_a(k) = 1\}|I_a^k)$ where $I_a^k := (\underline{X}_a^k, K_a^k)$ and we use $\underline{p}(k)$ to denote the $L \times 1$ vector representing the beliefs about all the channels in \mathcal{C} . The initial values of the belief parameters for all channels are set using the stationary distribution of the Markov chain. Now, using P to represent the transition probability matrix for the state transitions of each channel we define:

$$q_a(k) = P(1, 1)p_a(k-1) + P(0, 1)(1 - p_a(k-1)) \quad (6)$$

This $q_a(k)$ represents the belief of the state of channel a in slot k conditioned on past observations up to slot $k-1$. Then the belief values get updated as follows after the observation in time slot k :

$$p_a(k) = \frac{q_a(k)f_1(\underline{X}_a(k))}{q_a(k)f_1(\underline{X}_a(k)) + (1 - q_a(k))f_0(\underline{X}_a(k))} \quad (7)$$

when channel a was selected in slot k (i.e. $u_k = a$) or $p_a(k) = q_a(k)$ otherwise. Thus from (7) we see that updates for the sufficient statistic can be done using only the joint log-likelihood ratio of the observations, $\mathcal{L}(\underline{X}_a(k))$, instead of the entire vector of observations. Furthermore, from (2) we also see that the access decisions also depend only on the log-likelihood ratios. Hence we can conclude that this problem with vector observations is equivalent to one with scalar observations where the scalars represent the joint log-likelihood ratio of the observations of all the cooperating secondary users. Therefore, in the rest of this paper, we use a scalar observation model with the observation made on channel a at time k represented by $Y_a(k)$. We use Y^k to denote the set of all observations up to time slot k and Y_a^k to denote the set of all observations on channel a up to slot k . Hence the new optimal access decisions are decided by comparing the log-likelihood ratio of $Y_a(k)$ represented by $\mathcal{L}'(Y_a(k))$ to a threshold Δ' chosen such that:

$$P(\{\mathcal{L}'(Y_a(k)) < \Delta'\} | \{S_a(k) = 1\}) = \zeta \quad (8)$$

and access decisions are given by:

$$\delta_a(k) = \mathcal{I}_{\{\mathcal{L}'(Y_a(k)) < \Delta'\} \cap \{u_k = a\}} \quad (9)$$

Similarly the belief updates are done as in (7) with the evaluations of density functions of $\underline{X}_a(k)$ replaced with the evaluations of the density functions f'_0 and f'_1 of $Y_a(k)$. We use $G(\underline{p}(k-1), u_k, Y_{u_k}(k))$ to denote the function that returns the value of $\underline{p}(k)$ given that channel u_k was sensed in slot k . There is some randomness in function $G(\cdot)$ arising from the random observation $Y_{u_k}(k)$. The reward obtained in slot k can now be expressed as:

$$r_k = B(1 - \epsilon) \mathcal{I}_{\{S_{u_k}(k) = 0\}} \quad (10)$$

where ϵ is given by

$$\epsilon = P(\{\mathcal{L}'(Y_a(k)) > \Delta'\} | \{S_a(k) = 0\}) \quad (11)$$

From the structure of the dynamic program, it can be shown that the optimal solution to this dynamic program can be obtained by solving the following Bellman equation [6] for the optimal reward-to-go function:

$$J(\underline{p}) = \max_{u \in \mathcal{C}} [B(1 - \epsilon)(1 - q_u) + \alpha E(J(G(\underline{p}, u, Y_u)))] \quad (12)$$

where \underline{p} represents the initial value of the belief vector i.e., the prior probability of channel occupancies in slot -1 , and \underline{q} is calculated from \underline{p} as in (6). The expectation in (12) is performed over the random observation Y_u . Since it is not easy to find the optimal solution to this POMDP problem, we adopt a suboptimal strategy to get a channel selection policy that performs well.

In order to ensure that the resultant Markov chain is irreducible and positive recurrent we make the following assumption on the probability transition matrix P :

$$\text{Assumption : } 0 < P(j, j) < 1, j \in \{0, 1\} \quad (13)$$

A. Greedy policy

A straightforward strategy for tackling the channel selection problem is to employ the policy of maximizing the expected instantaneous reward in the current time slot. The expected instantaneous reward obtained by accessing some channel a in some slot k is given by $B(1 - \epsilon)(1 - q_a(k))$ where ϵ is given by (11). Hence the greedy policy would be to choose the channel a such that $1 - q_a(k)$ is the maximum:

$$u_k^{\text{gr}} = \operatorname{argmax}_{u \in \mathcal{C}} \{1 - q_u(k)\} \quad (14)$$

In other words, in every slot the greedy policy chooses the channel that is most likely to be free conditioned on the past observations. The greedy policy, besides being a simple policy to implement, has also some further justification. As we elaborate in [10], the results of [2] and [3] can be used to argue that the greedy policy is optimal under high SNR in many scenarios. The greedy policy for this problem is also equivalent to the Q_{MDP} policy which is a standard sub-optimal solution to the POMDP problem (see e.g., [7]).

B. An upper bound

A standard method for obtaining an upper bound on the reward of the optimal solution to a POMDP is by making the Q_{MDP} assumption [7], i.e., by assuming that in all future slots, the state of all channels become known exactly after making the observation in that slot. Under this assumption (12) can be solved exactly as we show in [10]. Since we are assuming more information than the maximum that we can obtain, the reward function under this assumption yields an analytic upper bound on the optimal reward of the original problem (12).

IV. THE CASE OF UNKNOWN DISTRIBUTIONS

In practice, the secondary users are typically unaware of the primary's signal characteristics and the channel realization from the primary [9], and hence have to rely on some form of non-coherent detection such as energy detection while sensing the primary signals. Furthermore, even while employing non-coherent detectors, the secondary users are also unaware of their locations relative to the primary and hence are not aware of the shadowing and path loss from the primary. Hence it is not reasonable to assume that the secondary users know the exact distributions of the observations under the primary-present hypothesis although it can be assumed that the distribution of the observations under the primary-absent hypothesis are known exactly. This scenario can be modeled by using a parametric description for the distributions of the received signal under the primary present hypothesis. We denote the density functions of the observations under the new hypotheses as:

$$\begin{aligned} H_0 : S_a(k) = 0 & : Y_a(k) \sim f_{\theta_0} \\ H_1 : S_a(k) = 1 & : Y_a(k) \sim f_{\theta_a} \end{aligned} \quad \text{where } \theta_a \in \Theta \subset \mathbb{R}, \quad a \in \mathcal{C} \quad (15)$$

where the parameters $\theta_a \in \Theta$ are unknown for all channels a , and $\theta_0 \in \mathbb{R}$ is known. We use $\mathcal{L}_{\theta}(\cdot)$ to denote the log-likelihood function under f_{θ} defined by: $\mathcal{L}_{\theta}(x) :=$

$\log\left(\frac{f_\theta(x)}{f_{\theta_0}(x)}\right)$. In this section, we study two possible approaches for dealing with such a scenario under specific assumptions about the parametric family. We restrict ourselves to greedy policies for channel selection, and for ease of illustration, consider a secondary system comprised of a single user.

A. Worst-case design for non-random θ_a

A close examination of section III reveals two specific uses for the density function of the observations under the H_1 hypothesis. The knowledge of this density was required for setting the access threshold in (8) to ensure the constraint on the probability of interference. When the parameters θ_a are non-random and unknown, there is very little that we can do in meeting the constraint on the interference probability. We will have to guarantee the constraint for all possible realizations of θ_a . The optimal access decision would thus be given by:

$$\hat{\delta}_a(k) = \mathcal{I}_{\{u_k=a\} \cap (\cap_{\theta \in \Theta} \{\mathcal{L}_\theta(Y_a(k)) < \tau_\theta\})} \quad (16)$$

where τ_θ satisfies:

$$P(\{\mathcal{L}_\theta(Y_a(k)) < \tau_\theta\} | \{S_a(k) = 1, \theta_a = \theta\}) = \zeta \quad (17)$$

The other place where this density was used was in updating the belief probabilities. Since θ_a are unknown, we have to choose some distribution for the observations under H_1 in order to perform the belief updates. We would ideally like to use the distribution corresponding to some worst-case value, $\theta^* \in \Theta$, such that our learning algorithm for tracking the channel states would give a better performance when the actual realization of θ_a is in fact different. This is possible if the densities described in (15) satisfy certain conditions [10]. If these conditions hold, a good sub-optimal solution to the channel selection problem would be to run the greedy policy for channel selection using f_{θ^*} for the density under H_1 while performing the updates of the channel beliefs.

B. Modeling θ_a as random

In section V-B, we show through simulations that the worst-case approach of the previous section leads to a severe decline in performance relative to the scenario where the distribution parameters in (15) are known accurately. In practice it may be possible to learn the value of these parameters online. In order to learn the parameters θ_a we need to have a statistical model for these parameters and a reliable statistical model for the channel state transitions. In this section we model the parameters θ_a as random variables that are i.i.d. across the channels and independent of the Markov process as well as the noise process. Unlike in the previous sections, we now assume that the Markov channel statistics are reliable, since they will be crucial in learning the parameters θ_a . Further, we assume that the cardinality of set Θ is finite³ and let $|\Theta| = N$. Let $\{\theta^i\}_1^N$ denote the elements of set Θ . The prior distribution of the parameter θ_a is known to the secondary users. The vector of beliefs of the different channels no longer forms a sufficient statistic for this problem. Instead, we keep track of the set of

³When Θ is a compact set an approximation of this algorithm can be used on a finite subset of Θ

joint a posteriori probabilities of the channel states and θ_a parameters, referred to as *joint beliefs*. We store these joint beliefs in $L \times N \times 2$ array $Q(k)$ with elements given by:

$$Q_{a,i,j}(k) = P(\{(\theta_a, S_a(k)) = (\theta^i, j)\} | I_a^k) \quad (18)$$

The a posteriori probabilities of θ_a parameters, which we refer to as the belief about θ_a , can be calculated from the array Q by just summing over the state indices.

The details of our sensing and access algorithms can be found in [10] which we do not include here due to lack of space. The basic idea is as follows. For each slot k , we identify a subset $\Theta_a(k)$ of Θ that contains all elements in Θ which have a high value for the belief about θ_a . We then design the access policy following an approach similar to that in (16), by ensuring that the interference constraint is satisfied conditioned on θ_a taking any value in $\Theta_a(k)$. This ensures that the overall probability of interference satisfies the interference constraint. Once the access policy is fixed, we design the greedy channel selection policy using an approach similar to that in (14) such that the new instantaneous reward is maximized. We also show that the belief about the θ_a parameters converge to 1 for the true value of the parameter in Θ and 0 for others. Thus, our algorithm asymptotically satisfies the interference constraint even conditioned on θ_a taking its true value.

V. SIMULATION RESULTS: AN EXAMPLE, AND COMPARISONS

For the simulations, we use the following parameters: $B = 1$, $L = 2$, $\alpha = 0.999$, $\zeta = 0.01$ and $P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$. We used the following distributions for $Y_a(k)$ under the two hypotheses:

$$\begin{aligned} H_0 : S_a(k) = 0 & : Y_a(k) = n_a(k) \\ H_1 : S_a(k) = 1 & : Y_a(k) = \theta_a + n_a(k) \\ \text{where } n_a(k) & \sim \mathcal{N}(0, \sigma^2), \theta_a \in \Theta \end{aligned} \quad (19)$$

where parameter θ_a can be interpreted as the mean power if the secondaries employ energy detection. This parameter may be unknown if the secondaries are unaware of their location relative to the primary transmitter.

A. Known θ_a : comparison with ACK-based scheme

For known θ_a , we use $\Theta = \{\mu\}$ in (19). We compared our greedy solution of section III-A that uses observations with the approach in [4] where error-free ACK signals from the secondary receivers are used as in the POMDP. We use $\sigma = 1$ and choose the values of μ such that $\text{SNR} = 20 \log_{10}(\mu/\sigma)$ takes values in the range 1dB - 10dB. For the distributions in (19), the access decision of (9) can be expressed as:

$$\delta_a(k) = \mathcal{I}_{\{Y_a(k) < \tau\} \cap \{u_k=a\}} \quad (20)$$

where τ is chosen such that:

$$P(\{Y_a(k) < \tau\} | \{S_a(k) = 1\}) = \zeta \quad (21)$$

The simulation results shown in Fig. 1 clearly show that using observations (G_1) is significantly better than relying on ACK

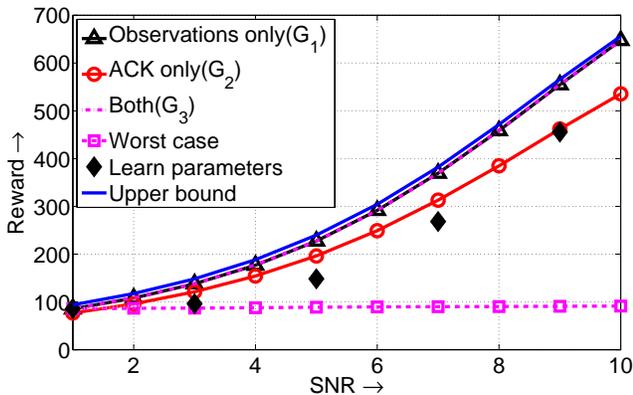


Fig. 1. Comparison of performances obtained

signals (G_2). Incorporating ACK signals in addition to the observations gives no significant advantage as seen in the overlapped curves of G_1 and G_3 . The explanation for this phenomenon is that for low values of ζ , the secondary users access some channel a in slot k only if the value of $Y_a(k)$ is small. Hence for a successful transmission to occur we should have $S_a(k) = 0$ and $Y_a(k)$ should be small. Thus ACK signals are received only when $Y_a(k)$ is small, in which case the observations themselves carry most of the information about the state of the channel, and hence little additional information is obtained from the ACK signals.

B. Unknown θ_a

Here, we compare the performances of both schemes of section IV for the hypotheses described in (19). We chose the set Θ such that the SNR values in dB given by $20 \log \frac{\theta_i}{\sigma}$ belong to the set $\{1, 3, 5, 7, 9\}$. The prior probability distribution for θ_a was chosen to be the uniform distribution on Θ . Both channels were assumed to have the same SNR values while the simulations were performed.

The results of these simulations are also given in Fig. 1. The net reward values obtained under the worst-case design of section IV-A and that obtained with the algorithm for learning θ_a given in section IV-B are shown. Clearly, we see that the worst-case design gives us almost no improvement in performance for high values of actual SNR. This is because the threshold we choose is too conservative for high SNR scenarios leading to several missed opportunities for transmitting. The minimal improvement in performance at high SNR is due to the fact that the system now has more accurate estimates of the channel beliefs although the update equations were designed for a lower SNR level. The learning scheme that uses a prior distribution on θ_a , on the other hand, yields a significant performance advantage over the worst-case scheme for high SNR values as seen in Fig. 1. We also see in the figure that the performance of the learning scheme with unknown θ_a relative to that of the greedy policy with known θ_a , improves at high SNR values. This is because learning of parameters is easier when the SNR is high.

VI. CONCLUSIONS

The main advantage that the ACK-based scheme of [4] has over our scheme of using observations, is that it allows synchronization between the secondary transmitter and receiver. However, in the presence of hidden terminals [8], the assumption of error-free ACKs fails to hold and this advantage is also lost. Thus we believe that a more practical solution to this synchronization problem would be to set aside a dedicated control channel for the purpose of reliably maintaining synchronization and make better use of the unused licensed spectrum by exploiting observations for tracking channel occupancies as illustrated by the results of section V-A.

From the results of section V-B, we conclude that the popular approach of designing for the worst-case value of the unknown mean power leads to a severe loss in performance when the actual SNR is high. The proposed learning-based scheme provides a substantial improvement in throughput over the worst case approach, especially at high SNR.

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 (JSAC): Special Issue on Adaptive, Spectrum Agile and Cognitive Wireless Networks*, vol. 25, no. 3, pp. 589-600, April 2007.
- [2] Q. Zhao, B. Krishnamachari, and K. Liu, "On Myopic Sensing for Multi-Channel Opportunistic Access: Structure, Optimality, and Performance," to appear in *IEEE Transactions on Wireless Communications*, available online at <http://www.ece.ucdavis.edu/~qzha/Journal.html>.
- [3] S. H. A. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of Myopic Sensing in Multi-Channel Opportunistic Access," submitted to *IEEE Transactions on Information Theory*, May 2008, available online at <http://www.ece.ucdavis.edu/~qzha/Journal.html>.
- [4] 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*, November 2007.
- [5] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.
- [6] D. Bertsekas, *Dynamic Programming and Optimal Control*, vol.1, Athena Scientific 2005.
- [7] D. Aberdeen, "A (revised) survey of approximate methods for solving POMDPs," Technical Report, Dec. 2003, <http://users.rsise.anu.edu.au/~daa/papers.html>.
- [8] J. Unnikrishnan, and V. V. Veeravalli, "Cooperative Sensing for Primary Detection in Cognitive Radio," *IEEE Journal of Selected Topics in Signal Processing, Special Issue on Signal Processing and Networking for Dynamic Spectrum Access*, vol. 2, pp. 18-27, February 2008.
- [9] A. Sahai, N. Hoven, and R. Tandra, "Some Fundamental Limits on Cognitive Radio," in *Forty-Second Allerton Conference on Communication, Control, and Computing*, Oct. 2004.
- [10] J. Unnikrishnan, and V. V. Veeravalli, "Algorithms for Dynamic Spectrum Access with Learning for Cognitive Radio," submitted to *IEEE Transactions on Signal Processing*, June 2008, available online at <http://www.ifp.uiuc.edu/~junnikr2/>.