Low-Complexity Approaches to Spectrum Opportunity Tracking

(Invited Paper)

Qing Zhao
University of California
Davis, CA 95616
Email: qzhao@ece.ucdavis.edu

Bhaskar Krishnamachari
University of Southern California
Los Angeles, CA 90089
Email: bkrishna@usc.edu

Keqin Liu
University of California
Davis, CA 95616
Email: kqliu@ucdavis.edu

***Abstract*—We consider opportunistic spectrum access under design constraints imposed at both node and link levels. First, hardware and energy limitations at node level may prevent a secondary user from sensing all the channels in the spectrum simultaneously. A channel selection strategy is thus necessary to track the time-varying spectrum opportunities. Second, sensing errors are inevitable. A secondary user needs to decide, based on imperfect sensing outcomes, whether to access the sensed channel and how to update its statistical knowledge of spectrum dynamics for better tracking in the future. Third, a secondary transmitter and its intended receiver need to hop synchronously in the spectrum in order to communicate. When a dynamic opportunity tracking strategy is used where the channel selection depends on the sensing history, achieving this synchrony is nontrivial in the absence of a dedicated control channel and in the presence of sensing errors. These practical constraints significantly complicate the design of opportunistic spectrum access, and the optimal performance requires the joint design of the spectrum sensor, opportunity tracking strategy, and spectrum access decisions. The focus of this paper is on developing low-complexity approaches for opportunistic spectrum access. We show that under certain conditions on the spectrum dynamics, simple myopic strategies can provide optimal performance for the joint design of spectrum sensor, opportunity tracking, and opportunity exploitation. We also propose an alternate low-complexity indexing strategy for other conditions that takes into account the expected time to channel availability.**

***Index Terms*—Opportunistic spectrum access, POMDP, myopic policy, spectrum opportunity tracking.**

I. Introduction

Various approaches to dynamic spectrum access have been envisioned to address the under-utilization of the radio spectrum as revealed by the measurements of actual spectrum usage [1]. Opportunistic spectrum access (OSA), also referred to as spectrum overlay, is perhaps the most compatible with the current spectrum management policy and the legacy systems. Built upon a hierarchical access structure, opportunistic spectrum access allows secondary users to exploit local and instantaneous spectrum availability provided that the interference to primary users is capped below a specified level.

A. Design Constraints in OSA

Several practical constraints at both node and link levels complicate the design of OSA. First, hardware and energy limitations at the node level may prevent a secondary user from sensing all the channels of interest simultaneously. A sensing strategy for intelligent channel selection is thus necessary to track the time-varying spectrum opportunities. The key to an efficient sensing strategy is the optimal exploitation of the entire observation history. This is because sensing outcomes provide statistical knowledge of the dynamics of spectrum opportunities and can thus guide the secondary user to channels likely to be idle. Sensing, as a consequence, has two functions: identifying an opportunity for immediate access, and gaining statistical information on channel usage for future use. The optimal sensing strategy needs to balance both. The design of such a strategy is, in general, a sequential decision making problem, where the current channel selection depends on specific realizations of the observation history. A predetermined channel selection sequence may lead to significant performance loss.

Second, sensing errors are inevitable in the wireless communication environment. An idle channel may be sensed as busy and *vice versa*. The former gives rise to a false alarm, and the latter to a miss detection. Given a potentially erroneous sensing outcome, the secondary user needs to decide whether to access. The tradeoff here is between minimizing overlooked spectrum opportunities and complying with the interference constraint. Clearly, the operating characteristics of the spectrum sensor (probability of false alarm vs. probability of miss detection) play an important role in access decision making. Sensing errors also complicate the extraction of statistical information on spectrum dynamics from sensing outcomes. Again, the operating characteristics of the spectrum sensor need to be taken into consideration.

The third design constraint is at the link level. A secondary transmitter and its intended receiver need to select the same channel in order to communicate with each other. Without a dedicated control channel to coordinate, achieving synchronous hopping in the spectrum is nontrivial unless a predetermined sequence of channel selections is used. With a dynamic sensing strategy, this synchrony requires that channel selections are based on an observation history common to both the transmitter and the receiver. Due to the random occurrence of sensing errors, channel selections at the transmitter and the receiver cannot be based on their individual sensing outcomes. A common observation is the history of acknowledgements immediately following successful transmissions, as considered in [2], [3]. In this case, an ACK tells unambiguously the true state of the sensed channel, while the absence of an ACK (or a NAK) may result from a no-access decision at the transmitter or a collision with primary users. This asymmetry has important implications in the access decision making and the update of spectrum statistical information for tracking purposes.

B. Scope and Focus

The above discussed constraints dictate the interaction across the sensing strategy, the spectrum sensor, and the access strategy. A joint design of these three basic components of OSA is thus necessary to achieve the optimal performance. Based on a Markovian model of the primary users' spectrum usage, the joint design of OSA has been formulated as a constrained partially observable Markov decision process (POMDP) in [3]. While powerful in modeling, POMDP suffers from the curse of dimensionality and seldom admits tractable solutions. Constraints on a POMDP further complicate the problem, often demanding *randomized* policies to achieve optimality.

Fortunately, it has been shown in [3] that there exists a separation principle in the joint design of OSA. Specifically, the optimal joint design can be carried out in two steps: first to choose the spectrum sensor and the access strategy to maximize the instantaneous throughput under a collision constraint, and then to choose the sensing strategy to maximize the overall throughput. This separation principle reveals the optimality of myopic policies for the design of the spectrum sensor and the access strategy, leading to closed-form optimal solutions. Furthermore, decoupling the design of the sensing strategy from that of the spectrum sensor and the access strategy, the separation principle reduces the constrained POMDP to an unconstrained one, which admits deterministic optimal policies.

This result shows that the sensing strategy for opportunity tracking can be designed separately from the spectrum sensor and the access strategy. In the design of the spectrum sensor and the access strategy, the myopic approach is optimal, which reduces a sequential decision making problem to a static optimization and leads to the optimal solution in closed-form.

The complexity issue in the joint design of OSA is, however, only partially resolved. Designing the optimal sensing strategy remains to be a full-blown POMDP. The focus of this paper is on developing low-complexity sensing strategies and studying their optimality. Together with the results given in [3], we hope to provide a more complete picture of the tradeoff between optimality and complexity in the design of OSA under practical constraints. Specifically, we show that under certain conditions on the Markovian model of spectrum occupancy and the operating statistics of the spectrum sensor, the myopic approach is also optimal in the design of sensing strategies. This result complements the separation principle that has revealed the optimality of the myopic approach in the design of the spectrum sensor and access strategy. We further show that the myopic sensing strategy has a simple structure; optimal channel selections can be reduced to a counting procedure.

We also propose another low-complexity indexing strategy that takes into account the expected time till a channel becomes free. We find that in some cases where the myopic policy is sub-optimal, this alternate approach can provide better performance.

This paper builds upon our previous results on the optimality of myopic sensing in the absence of sensing errors [4]. Sensing errors, however, significantly complicate opportunity tracking due to the constraint on transceiver synchrony.

C. Related Work

The literature on dynamic spectrum access is fast growing. An overview of challenges and recent developments can be found in [1]. In the context of networking protocol design for OSA, many existing results assume perfect full spectrum sensing. Given the design constraints discussed in Section 1.1, efficient spectrum opportunity tracking becomes both necessary (due to partial spectrum sensing) and challenging (due to sensing errors and the need of transceiver synchrony).

OSA with partial yet perfect spectrum sensing has been addressed in [5], [6]. In [5], MAC protocols are proposed for an ad hoc secondary network overlaying a GSM cellular network. It is assumed that the secondary transmitter and receiver exchange information on which channel to use through a commonly agreed control channel. In [6], access strategies for slotted secondary users searching for opportunities in an un-slotted primary network is considered under a continuous-time Markovian model of channel occupancy. A round-robin single-channel sensing scheme is used and is shown to be near optimal when the interference constraint is strict. The POMDP framework for the joint design of OSA under partial imperfect sensing was proposed and studied in [2], [3]. An overview of this framework can be found in [7].

Crucial to the efficiency of spectrum opportunity tracking are simple yet sufficiently accurate statistical models of spectrum occupancy. Several testbeds have been established to monitor the actual spectrum usage in different frequency bands [8]–[10], and Markovian and semi-Markovian models have been shown to fit well with spectrum usage measurements. With these active experimental research activities, we can perhaps foresee a public database of statistical models of spectrum usage in different bands and at different time and location. Secondary users can then download the required model for the design of spectrum sensing and access strategies.

II. The Network Model

Consider a spectrum consisting of N channels, each with bandwidth B_i $(i = 1, \cdots, N)$. These N channels are licensed to a primary network whose users communicate according to a synchronous slot structure. We model the spectrum occupancy as a discrete-time homogenous Markov process with 2^N states. A state $\mathbf{S}(t) \stackrel{\Delta}{=} [S_1(t), \ldots, S_N(t)]$ denotes the channel occupancy, where $S_n(t) \in \{0 \text{ (busy)}, 1 \text{ (idle)}\}$ is the state of channel n in slot t.

We consider a secondary network whose users independently and selfishly search for and access spectrum opportunities in these N channels. In each slot, a secondary user chooses one of the N channels to sense. Based on the sensing outcome which is subject to errors, the user decides whether to access the channel. Accessing an idle channel leads to bit delivery while accessing a busy channel results in a collision with the primary network. At the end of the slot, the receiver acknowledges a successful transmission.

The objective is to design jointly the spectrum sensor, the sensing strategy for channel selection, and the access strategy determining whether to transmit so that the throughput of the secondary user is maximized under the constraint that the probability of colliding with primary users is capped below ζ in any channel and slot.

We make the following assumptions.

A1: The transition probabilities of the Markovian model remain unchanged for T slots and are known to the secondary users.

A2: Acknowledgements are received without error (note that acknowledgements are always transmitted over an available channel).

III. Optimal Joint Design of OSA

In this section, we review the result obtained in [3], where the joint design of OSA under partial imperfect sensing is formulated as a constrained POMDP. A separation principle was established, decoupling the sensing strategy from the spectrum sensor and the access strategy in the joint design and providing a closed-form solution to the optimal design of the latter two components. The explicit characterization of the optimal spectrum sensor and access strategy significantly simplifies the study of low complexity sensing strategies as addressed in Section 4.

A. Spectrum Sensor

The question we aim to answer in designing the spectrum sensor is which criterion should be adopted, the Bayes or the Neyman-Pearson (NP). If the former, how do we choose the risks? If the latter, how should we set the constraint on the probability of false alarm?

Suppose that channel n is chosen in slot t. The spectrum sensor detects the presence of primary users in this channel by performing a binary hypothesis test:

$$\mathcal{H}_0 : S_n(t) = 1 \text{ (idle)} \quad \text{vs.} \quad \mathcal{H}_1 : S_n(t) = 0 \text{ (busy)}. \qquad (1)$$

The performance of the spectrum sensor is characterized by the probability of false alarm (PFA) $\epsilon(t)$ and the probability of miss detection (PM) $\delta(t)$:

$$\begin{aligned} \epsilon(t) &\triangleq \Pr\{\text{decide } \mathcal{H}_1 \,|\, \mathcal{H}_0 \text{ is true}\}, \\ \delta(t) &\triangleq \Pr\{\text{decide } \mathcal{H}_0 \,|\, \mathcal{H}_1 \text{ is true}\}. \end{aligned}$$

For a given PFA $\epsilon(t)$, the largest achievable probability of detection, denoted as $P_{D,\max}(\epsilon(t))$, can be attained by the optimal NP detector with the constraint that the PFA is no larger than $\epsilon(t)$, or an optimal Bayesian detector with a suitable set of risks [11, Sec. 2.2.1]. All operating points (ϵ, δ) above the best ROC curve $P_{D,\max}$ are thus infeasible. An illustration of all feasible operating points is given in Fig. 1. We also note that every feasible sensor operating point (ϵ, δ) lies on a line that connects two boundary points and hence can be achieved by randomizing between two optimal NP detectors with properly chosen constraints on the PFA [11, Sec. 2.2.2]. Therefore, the design of spectrum sensor is reduced to the choice of a desired sensor operating point.

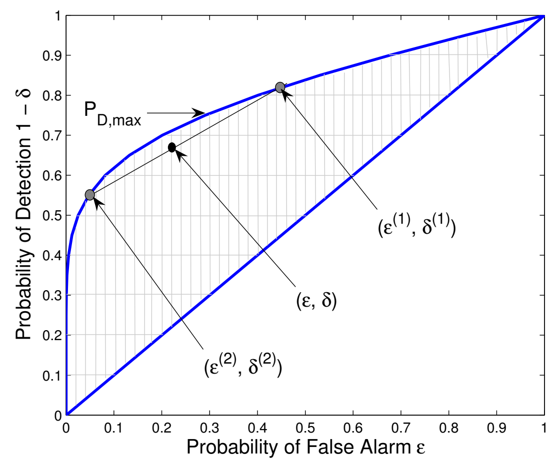

Fig. 1. Illustration of the set of all feasible sensor operating points (ϵ, δ).

The design of the optimal NP detector is a well-studied classic problem, which is not the focus here. Our objective is to define the criterion and the constraint under which the spectrum sensor should be designed, equivalently, to find the optimal sensor operating point $(\epsilon^*(t), \delta^*(t))$ to achieve the best tradeoff between false alarm and miss detection in each slot t.

B. Sensing and Access Strategies

In each slot, a sensing strategy decides which channel in the spectrum to sense, and an access strategy determines whether to access given the sensing outcome. Below we illustrate the sequence of operations in each slot.

At the beginning of slot t, the system state transits to $\mathbf{S}(t) = [S_1(t), \ldots, S_N(t)]$ according to the transition probabilities of the underlying Markov process. The secondary user first chooses a channel $a(t)$ to sense and a feasible sensor operating point $(\epsilon(t), \delta(t))$. It then decides whether to access

$$\Phi_a(t) \in \{0 \text{ (no access)}, 1 \text{ (access)}\}$$

by taking into account the sensing outcome provided by the spectrum sensor that is designed according to the chosen operating point. A collision with primary users happens when the secondary user accesses a busy channel. At the end of this slot, the receiver acknowledges a successful transmission $K_a(t) \in \{0 \text{ (no ACK)}, 1 \text{ (ACK)}\}$.

C. Constrained POMDP Formulation

Due to partial spectrum monitoring and sensing errors, the internal state of the underlying Markov process that models spectrum occupancy cannot be fully observed. Considering the constraint on the collision probability, we can formulate the joint design of OSA as a constrained POMDP over a finite horizon of length T.

Reward The reward can be defined as the number of bits delivered, which is assumed to be proportional to the channel bandwidth. Given sensing action $a(t)$ and access action $\Phi_a(t)$, the immediate reward $R(t)$ in slot t is given by

$$R(t) = S_a(t)\Phi_a(t)B_a. \tag{2}$$

Hence, the expected total reward of the POMDP represents overall throughput, the expected total number of bits that can be delivered by the secondary user in T slots.

Belief Vector Due to partial spectrum monitoring and sensing errors, a secondary user cannot directly observe the true system state. It can, however, infer the SOS from its decision and observation history. As shown in [12], the statistical information on the system state provided by the entire decision and observation history can be encapsulated in a belief vector $\mathbf{\Lambda}(t) \stackrel{\Delta}{=} \{\lambda_{\mathbf{s}}(t)\}_{\mathbf{s}}$, where $\lambda_{\mathbf{s}}(t)$ denotes the conditional probability (given the decision and observation history) that the system state is $\mathbf{s}$ at the beginning of slot t *prior to* the state transition.

To ensure synchronous hopping in the spectrum without introducing extra control message exchange, the secondary user and its desired receiver must have the same history of observations so that they make the same channel selection decisions. Since sensing errors may cause different sensing outcomes at the transmitter and the receiver, the acknowledgement $K_a(t) \in \{0, 1\}$ should be used as the common observation in each slot.

Policy A joint design of OSA is given by policies of the above POMDP. Specifically, a sensing policy π_s specifies a sequence of functions, each mapping a belief vector $\mathbf{\Lambda}(t)$ at the beginning of slot t to a channel $a(t)$ to be sensed in this slot. For a finite-horizon POMDP, the optimal mapping is, in general, time-varying, *i.e.,* the optimal policy is non-stationary. Similarly, a sensor operating policy π_δ specifies, in each slot t, a spectrum sensor design $(\epsilon(t), \delta(t))$ based on the current belief vector $\mathbf{\Lambda}(t)$. An access policy π_c specifies an access decision $\Phi_a(t)$ based on the current belief vector $\mathbf{\Lambda}(t)$ and the sensing outcome.

The above defined policies are deterministic. For constrained POMDPs, we may need to resort to randomized policies to achieve optimality. In this case, what needs to be chosen is not specific actions, but the probability of taken each possible action. Due to the uncountable space of probability distributions, randomized policies are usually computationally prohibitive.

Optimal Joint Design The optimal joint design of OSA is then given by the following constrained POMDP.

$$\begin{aligned} &\{\pi_\delta^*, \pi_s^*, \pi_c^*\} = \arg\max_{\pi_\delta, \pi_s, \pi_c} \mathbb{E}_{\{\pi_\delta, \pi_s, \pi_c\}} \left[\sum_{t=1}^{T} R(t) \middle| \mathbf{\Lambda}(1) \right] \\ &\text{s.t. } P_a(t) = \Pr\{\Phi_a(t) = 1 \,|\, S_a(t) = 0\} \leq \zeta, \qquad \forall a, t, \end{aligned} \tag{3}$$

where $\mathbb{E}_{\{\pi_\delta, \pi_s, \pi_c\}}$ represents the expectation given that policies $\{\pi_s, \pi_\delta, \pi_c\}$ are employed, $P_a(t)$ is the probability of collision perceived by the primary network in chosen channel $a(t)$ and slot t, and $\mathbf{\Lambda}(1)$ is the initial belief vector, which can be set to the stationary distribution of the underlying Markov process if no information on the initial system state is available.

D. The Separation Principle

A separation principle for the joint design of OSA has been established in [3]. It is shown that the joint design can be carried out in two steps without losing optimality: (i) obtain the optimal sensor operating policy π_δ^* and the optimal access policy π_c^* by maximizing the instantaneous reward $R(t)$ in the current slot under the collision constraint; (ii) obtain the optimal sensing policy π_s^* to maximize the overall throughput using π_δ^* and π_c^* obtained in the first step.

The separation principle decouples the design of the sensing policy from that of the spectrum sensor and access policy. As a consequence, the design of the sensing policy is reduced to an *unconstrained* POMDP, where optimality is achieved with *deterministic* policies. Furthermore, it reveals that the optimal sensor operating policy π_δ^* and the optimal access policy π_c^* can be obtained from a myopic approach that focuses solely on the instantaneous reward and ignores the impact of the current actions on the future reward. The joint design of π_δ^* and π_c^* is thus reduced to a static optimization problem with a simple, time-invariant, and closed-form solution. Specifically, the optimal sensor should adopt the optimal NP detector with constraint $\delta_a^* = \zeta$ on the probability of miss detection. Correspondingly, the optimal access policy is to trust the sensing outcome given by the spectrum sensor, *i.e.,* access if and only if the channel is sensed to be idle.

What remains to be derived are low complexity sensing strategies in order to achieve tractable solutions to the joint design of OSA. This is the focus of the subsequent sections, where we study simple index policies and identify conditions under which they offer optimal performance.

IV. Index Sensing Policy: The Myopic Approach

Using the optimal sensor operating policy π_δ^* and the optimal access policy π_c^*, we can obtain the optimal sensing policy π_s^*, we can obtain the optimal sensing policy by solving an unconstrained POMDP.

$$\pi_s^* = \arg\max_{\pi_s} \mathbb{E}_{\{\pi_s\}} \left[\sum_{t=1}^{T} R(t) \middle| \mathbf{\Lambda}(1) \right].$$

Let $V_t(\mathbf{\Lambda}(t))$ denote the maximum expected remaining reward that can be accrued starting from slot t when the current belief vector is $\mathbf{\Lambda}(t)$. We have the following optimality equation.

$$V_t(\mathbf{\Lambda}(t)) = \max_{a=1,\cdots,N} \mathbb{E}[R_a(t) + V_{t+1}(\mathcal{T}(\mathbf{\Lambda}(t)|a, K_a(t)))], \quad (4)$$

where $R_a(t)$ is the immediate reward in slot t for action a, and $\Lambda(t+1) = \mathcal{T}(\mathbf{\Lambda}(t)|a, K_a(t))$ is the updated belief vector based on the action a and observation $K_a(t)$, which can be obtained via the Bayes rule as shown in Section 4.1.

Solving for the optimal policy from (4) has a complexity that grows exponentially with the horizon length T. Next, we seek low complexity sensing policies. We show that under certain conditions, low complexity policies with a simple structure can achieve the optimal performance.

A. Index Policy and Myopic Sensing

One family of low complexity policies for POMDPs is based on indexing the actions. Specifically, we assign an index $\gamma_i(\Lambda(t))$ to each action i and choose the action with the largest index determined by the current belief vector $\mathbf{\Lambda}(t)$. This family of policies are stationary: the mapping from the belief vector to the chosen action is the same for all slots. The only design parameter is the index function, which maps the belief vector to a real number.

The myopic policy is an index policy where the index of an action is the immediate reward obtainable through that action. Consider the spectrum opportunity tracking problem. Let $\mathbf{P}$ denote the transition matrix of the underlying Markov process. The distribution of the system state $\mathbf{S}(t)$ in slot t is then given by

$$\mathbf{\Lambda}'(t) = \mathbf{\Lambda}(t)\mathbf{P}.$$

Note that $\mathbf{\Lambda}(t)$ denotes the distribution of system state *prior to* the state transition at the beginning of slot t, a conventional definition in the literature of POMDP. From $\mathbf{\Lambda}'(t)$, we can then obtain the marginal probability $\omega_i'(t)$ that channel i is available in slot t. The index of the myopic sensing policy is thus simply

$$\gamma_i(\Lambda(t)) = \omega_i'(t)B_i. \quad (5)$$

At the end of slot t, we obtain the updated belief vector $\mathbf{\Lambda}(t+1)$ from the chosen action $a = \arg\max_i \gamma_i(\mathbf{\Lambda}(t))$ and the observation $K_a(t)$ using Bayes' rule as follows.

$$\lambda_{\mathbf{s}}(t+1) = \frac{\sum_{\mathbf{s}'} \lambda_{\mathbf{s}'}(t) P_{\mathbf{s}',\mathbf{s}} U_{\mathbf{s},k}(a)}{\sum_{\mathbf{s}} \sum_{\mathbf{s}'} \lambda_{\mathbf{s}'}(t) P_{\mathbf{s}',\mathbf{s}} U_{\mathbf{s},k}(a)}, \quad (6)$$

where $P_{\mathbf{s}',\mathbf{s}}$ is the transition probability from state $\mathbf{s}'$ to $\mathbf{s}$, and $U_{\mathbf{s},k}(a) \stackrel{\Delta}{=} \Pr\{K_a = k \,|\, \mathbf{S} = \mathbf{s}\}$ is the conditional distribution of the acknowledgement given current state $\mathbf{s}$ and action a. Recall that the optimal sensor operating point is $(\epsilon^*, \delta^* = \zeta)$ and the optimal access strategy is to transmit if and only if the channel is sensed to be idle. We can obtain $U_{\mathbf{s},k}(a)$ as follows.

$$U_{\mathbf{s},1}(a) \stackrel{\Delta}{=} \Pr\{K_a = 1 \,|\, \mathbf{S} = \mathbf{s}\} = s_a(1-\epsilon^*), \quad (7a)$$

$$U_{\mathbf{s},0}(a) = 1 - U_{\mathbf{s},1}(a), \quad (7b)$$

where s_a is the state of channel a (the ath component of $\mathbf{s}$). It is thus clear that the belief update for spectrum opportunity tracking depends on the sensor operating point and the access strategy. An explicite characterization of their optimal design simplifies the study of spectrum sensing strategies.

We show next that under certain conditions on the Markovian model and the false alarm probability ϵ^*, the myopic sensing policy has a simple structure and achieves the optimal performance. In this case, we do not even need to update the belief vector as given in (6) and (7). The optimal channel selection is reduced to a simple counting procedure, and low complexity is achieved without sacrificing performance.

B. The Structure and Optimality of Myopic Sensing for i.i.d. Channels

We consider the case that the channels evolve according to N independent and identical Markov processes. The state diagram and transition probabilities $\{p_{i,j}\}$ of each channel are illustrated in Figure 2. We show in this section that for i.i.d. channels, the myopic sensing policy has a simple structure. We further prove that the myopic sensing policy is optimal when $N = 2$. For $N > 2$, extensive numerical results have demonstrated the optimality of the myopic policy.

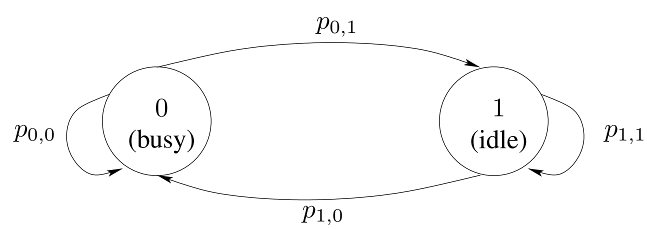

Fig. 2. The Markov channel model

Proposition 1: Consider N i.i.d. channels with $p_{0,1} > p_{1,1}$. Assume that the false alarm probability ϵ of the spectrum sensor satisfies

$$\epsilon^* < \frac{p_{0,0}p_{1,1}}{p_{0,1}p_{1,0}}.$$

In slot t, let $\tau_i(t) \in \{1, 2, \cdots, t-1, \infty\}$ denote the time difference between t and the last visit to channel i. If channel i has never been visited, then $\tau_i(t) = \infty$. Define the following sets.

$$\Delta_e(t) \;\stackrel{\Delta}{=}\; \{\tau_i(t) : \tau_i(t) \text{ is even}\},$$
$$\bar{\Delta}_e(t) \;\stackrel{\Delta}{=}\; \{\tau_i(t) : \tau_i(t) \text{ is odd or } \infty\}.$$

Given the action $a(t-1)$ and observation $K_{a(t-1)}(t-1)$ (acknowledgement) in slot $t-1$, the myopic action $a_*(t)$ in slot t that maximizes the expected immediate reward is as follows.

$$a_*(t) = \begin{cases} a(t-1) & \text{if } K_{a(t-1)}(t-1) = 0 \\ \arg\min \Delta_e(t) & \text{if } K_{a(t-1)}(t-1) = 1, \Delta_e(t) \neq \emptyset \\ \arg\max \bar{\Delta}_e(t) & \text{if } K_{a(t-1)}(t-1) = 1, \Delta_e(t) = \emptyset \end{cases}. \quad (8)$$

Proposition 1 shows that for i.i.d. channels with $p_{0,1} > p_{1,1}$, the optimal action under myopic sensing is to stay in the same channel after a NAK and switch to another channel after an

ACK, provided that the false alarm probability of the spectrum sensor is below a certain value. When a channel switch is needed, the user chooses, among those channels to which the last visit occurred an even number of slots ago, the one most recently visited. If there are no such channels, the user chooses the channel that has not been visited for the longest time, which can be any of the channels that have never been visited if such channels exist. For the case of $N = 2$, the myopic policy is simply to stay in the same channel after a NAK and switch to the other channel after an ACK.

The structure of myopic sensing for i.i.d. channels with $p_{0,1} < p_{1,1}$ can be similarly obtained, where the condition on the false alarm probability is given by

$$\epsilon^* < \frac{p_{1,0}p_{0,1}}{p_{1,1}p_{0,0}}.$$

In this case, the optimal action under myopic sensing is to stay in the same channel after an ACK and switch to the channel visited the longest time ago after a NAK.

We have assumed that no initial information on the system state is available in the first slot, *i.e.,* the initial distribution of the Markov chains is the stationary distribution. The myopic action in the first slot is to choose an arbitrary channel. It is straightforward to modify Proposition 1 when the initial distribution is not the stationary distribution.

The above proposition reveals that obtaining the myopic actions for i.i.d. channels is reduced to a simple counting procedure: the secondary user only needs to set up 4 pointers indicating the channels to which the last visits occurred most recently or the longest time ago (considering even and odd time differences separately). The complexity of obtaining the optimal myopic sensing policy is $\mathcal{O}(NT)$, linear in both N and T. We show in Theorem 1 below that the myopic sensing policy with such a simple structure is, in fact, optimal when $N = 2$.

Theorem 1: For two i.i.d. channels with $p_{0,1} > p_{1,1}$, the myopic sensing policy is optimal when the false alarm probability ϵ of the spectrum sensor satisfies

$$\epsilon^* < \frac{p_{0,0}p_{1,1}}{p_{0,1}p_{1,0}}.$$

Proof: The proof is based on the following lemma which applies to any POMDP over a finite horizon [4]. Details are given in Appendix II.

Lemma 1: Consider a general POMDP with a finite horizon of length T. A sufficient condition for the optimality of the myopic policy is given below.

C0: Among all actions in slot t ($t = 1, \cdots, T-1$), the myopic action maximizes the total expected remaining reward obtained by taking myopic actions in each of the remaining slots $t+1, \cdots, T$.

The optimality for two i.i.d. channels with $p_{0,1} < p_{1,1}$ can be similarly proven. Numerical results have demonstrated the optimality of the myopic sensing policy for $N > 2$. We are currently extending the proof of Theorem 1 to the general case. One numerical example is given in Figure 3, where we compare the throughput performance of the myopic sensing and the optimal policy. We consider 3 independent channels. In the upper figure, these channels are identical, while in the lower figures, channels have different transition matrixes. For spectrum opportunity detection, we assume that the background noise and the signal of primary users can be modeled as white Gaussian processes as considered in [2]. The maximum allowable collision probability is set to $\zeta = 0.05$. Based on the separation principle, the spectrum sensor operates at $\delta = 0.05$, which leads to a false alarm probability of $\epsilon = 0.0274$ for the chosen detection scenario. We observe that for i.i.d. channels, the performance of myopic sensing matches with the optimal performance. For nonidentical channels, there is performance loss. We point out that with both myopic sensing and the optimal sensing strategies, the throughput of the secondary user increases over time, which results from the improved information on the system state drawn from accumulating observations. This demonstrates the cognitive nature of these sensing strategies developed under the POMDP formulation: learning from and adapting to the communication environment for improved performance. The performance of the random channel selection scheme, however, remains the same over time.

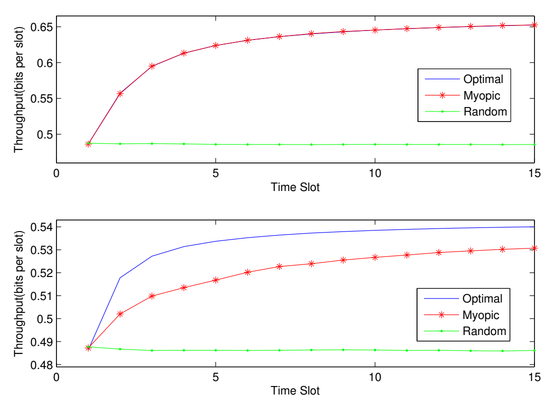

Fig. 3. Performance of myopic sensing.

V. A Suboptimal Strategy for Non-i.i.d. Channels

Figure 3 shows that for non-i.i.d. channels, myopic sensing results in performance loss as compared to the optimal policy. In this section, we consider other choices of the index function that may offer improved performance over the myopic approach for non-i.i.d. channels.

We define the index of a channel as the ratio of its bandwidth to the expected time until it becomes idle for the first time:

$$\gamma_i(\Lambda) = \frac{B_i}{\mathbb{E}[O_i(\Lambda)]},$$

where $O_i(\Lambda)$ denote the time until the first availability of channel i given the current belief vector Λ. In other words, this index policy chooses the channel that will become available most frequently (weighted by channel bandwidths).

We show next that the expected time until the first availability $\mathbb{E}[O_i(\Lambda)]$ can be easily obtained. Consider first that channels are independent but not necessarily identical. In this case, the marginal conditional distribution of channel states is a sufficient statistic [2], *i.e.,* we can consider the following belief vector

$$\Omega(t) = [\omega_1(t), \cdots, \omega_N(t)]$$

where $\omega_i(t)$ denotes the conditional probability that channel i is available at the beginning of slot t prior to the state transition. Note that the dimension of the belief vector is reduced from 2^N to N when channels are independent. Let

$$\omega_i' = \omega_i p_{1,1}^{(i)} + (1-\omega_i)p_{0,1}^{(i)}$$

denote the probability that channel i is available after the state transition at the beginning of the slot, where $\{p_{m,n}^{(i)}\}$ are the transition probabilities of channel i. We can obtain the distribution of $O_i(\Omega)$ as follows.

$$\Pr[O_i = k] = \begin{cases} \omega_i' & \text{if } k = 1 \\ (1-\omega_i')(p_{0,0}^{(i)})^{k-2}p_{0,1}^{(i)} & \text{if } k \geq 2 \end{cases}.$$

The index of channel i can then be obtained in closed-form:

$$\gamma_i(\Omega) = \frac{B_i p_{0,1}^{(i)}}{1 + p_{0,1}^{(i)} - \omega_i'}.$$

When channels are identical, the index is reduced to $\gamma_i(\Omega) = B_i\omega_i'$, which is the myopic approach. When channels are correlated, the expected time until the first availability $\mathbb{E}[O_i(\Lambda)]$ can be obtained by calculating the absorbing time of a Markov process.

Shown in Figure 4 is a simulation example where we compare the performance of this approach (MTTA: minimum time to availability) with the myopic policy. We observe that this index policy can offer improved performance over the myopic approach.

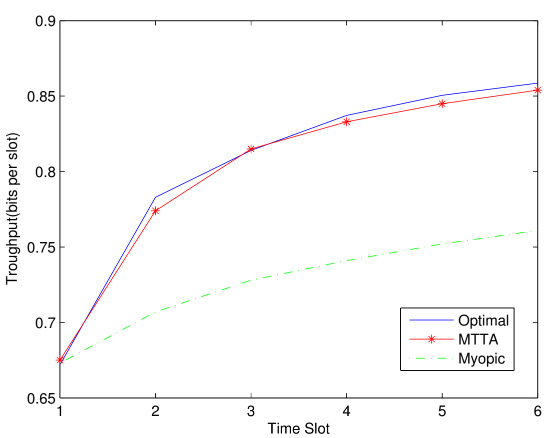

Fig. 4. Performance of an index policy based on the expected time to the first availability ($N = 7$).

We point out that when there are a small number of channels and the channels change state infrequently (for example, small $p_{0,1}$ and $p_{1,0}$ for independent channels), this index policy may perform worse than the myopic approach. It is perhaps not surprising that the optimal choice of the index function depends on the Markovian model, and certainly interesting to study the relationship between the Markovian statistics and the optimal index policy for spectrum opportunity tracking.

VI. Conclusion and Future Work

We have shown in this work that for the case of i.i.d. channels, the myopic sensing strategy has a simple structure. Furthermore, we have proven that for two i.i.d. channels, even in the presence of sensing errors, the myopic strategy provides optimal performance for the problem of spectrum opportunity tracking. This is a non-trivial extension of our previous results pertaining to the case of error-free sensing [4], as sensing errors make it challenging to maintain transmitter-receiver synchronization. Simulations indicate that this result holds generally for arbitrarily many channels. We are working to prove this analytically.

We also presented another index scheme, which ranks the channels at each step using the ratio of their bandwidth to the expected time to availability. In the case of non-i.i.d. channels, particularly when there are a large number of channels, this strategy appears to be a promising alternative to the myopic scheme. An interesting direction for future work is to identify the best indexing strategy for non-i.i.d. channels and characterize its performance with respect to the optimal.

Acknowledgement

This work was supported by the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011 and by the National Science Foundation under Grants CNS-0627090 and ECS-0622200.

APPENDIX I: PROOF OF PROPOSITION 1

Consider first $N = 2$. Without loss of generality, assume $a(t-1) = 1$. Consider first $K_1(t-1) = 1$. In this case, we know that $S_1(t-1) = 1$. The immediate reward for staying in channel 1 in slot t is $p_{1,1}B$, while the immediate reward for switching to channel 2 in slot t is

$$(\omega_2(t)p_{1,1} + (1-\omega_2(t))p_{0,1})B \geq p_{1,1}B, \quad \forall \omega_2(t) \in [0,1],$$

where the inequality follows from $p_{0,1} > p_{1,1}$. Hence, the myopic action in slot t is to switch to channel 2.

We now consider $K_1(t-1) = 0$. In this case, $S_1(t-1)$ can be either 0 or 1 due to sensing errors. Let $[\omega_1(t), \omega_2(t)]$ denote the belief vector at the beginning of slot t, *i.e.,* $\omega_i(t)$ is the probability that channel i is available at the beginning of slot t *prior* to the state transition, which is equivalent to the *a posterior* probability of $S_i(t-1) = 1$ after incorporating the observation obtained in slot $t-1$. Let $\omega_i'(t)$ be the probability that $S_i(t) = 1$, *i.e.,*

$$\omega_i'(t) = \omega_i(t)p_{1,1} + (1-\omega_i(t))p_{0,1} = p_{0,1} + \omega_i(t)(p_{1,1} - p_{0,1}).$$

We show below that $\omega_1'(t) \geq \omega_2'(t)$ when $K_1(t-1) = 0$, which implies that the myopic action in slot t is to stay in channel 1. Since $p_{0,1} > p_{1,1}$, it suffices to show $\omega_1(t) \leq \omega_2(t)$ (see the above equation), where

$$\begin{aligned}
\omega_1(t) &\triangleq \Pr[S_1(t-1)=1|K_1(t-1)=0]\\
&= \frac{\Pr[K_1(t-1)=0|S_1(t-1)=1]\Pr[S_1(t-1)=1]}{\Pr[K_1(t-1)=0]}\\
&= \frac{\epsilon\omega_1'(t-1)}{\epsilon\omega_1'(t-1)+(1-\omega_1'(t-1))},\\
\omega_2(t) &= \omega_2'(t-1).
\end{aligned}$$

Considering the condition on ϵ and the lower and upper bounds on $\omega_i'(t-1)$ obtained from $p_{0,1} > p_{1,1}$:

$$p_{1,1} \leq \omega_i'(t-1) \leq p_{0,1},$$

we arrive at $\omega_1(t) \leq \omega_2(t)$.

Applying the above arguments to $N > 2$, we know that the myopic action is to stay after observing $K = 0$ and switch after observing $K = 1$. The only question to address is which channel to switch to. The myopic action in slot t is to choose the channel that is most likely to be idle in slot t. Since we only switch channel after observing $K = 1$, the last known state of every channel is 1 (error free). The choice of channel is thus the same as in the perfect sensing case. The proof can be found in [4]

APPENDIX II: PROOF OF THEOREM 1

We first establish the following two lemmas.

Lemma 2: Under the conditions of Theorem 1, the expected total remaining reward starting from slot t under the myopic sensing policy is determined by the action $a(t-1)$ and the system state $\mathbf{S}(t-1)$ in slot $t-1$, hence independent of the belief vector $\Omega(t)$ at the beginning of slot t. Let $V_t(a(t-1), \mathbf{S}(t-1))$ denote the expected total remaining reward starting from slot t under the myopic sensing policy for given $a(t-1)$ and $\mathbf{S}(t-1)$. We further have

$$\begin{aligned}
V_t(1,[0,0]) &= V_t(2,[0,0])\\
V_t(1,[1,1]) &= V_t(2,[1,1])\\
V_t(1,[0,1]) &= V_t(2,[1,0])\\
V_t(1,[1,0]) &= V_t(2,[0,1]) \qquad (9)
\end{aligned}$$

Proof: Given $a(t-1)$ and $\mathbf{S}(t-1)$, the myopic actions in slots t to T, governed by Proposition 1, are fixed for each sample path of system state and observation, independent of $\Omega(t)$. As a consequence, the total reward obtained in slots t to T for each sample path is independent of $\Omega(t)$, so is the expected total reward. The four equalities can also be shown by considering each sample path of the system state and observations.

□□□

Note that Lemma 2 does not hold for a general POMDP, where the value function V_t of a given policy (expected total remaining reward starting from slot t) is determined by the belief vector $\Omega(t)$. For fixed action and observation in slot $t-1$, $\Omega(t)$ may be different for different $\Omega(t-1)$. In general, we should have

$$\begin{aligned}
&V_t(a(t-1), \mathbf{S}(t-1), \Omega(t-1))\\
=\ &\textstyle\sum_\theta \Pr[\theta|a(t-1), \mathbf{S}(t-1)] V_t(\mathcal{T}(\Omega(t-1)|a(t-1),\theta)),
\end{aligned}$$

where $\Pr[\theta|a(t-1), \mathbf{S}(t-1)]$ is the probability of observing θ in slot $t-1$ given $a(t-1)$ and $\mathbf{S}(t-1)$, and $\mathcal{T}(\Omega(t-1)|a(t-1),\theta) = \Omega(t)$ is the updated belief vector in slot t after incorporating $a(t-1)$ and θ.

□□□

Lemma 3: Under the conditions of Theorem 1, we have, $\forall t, a$,

$$|V_t(a,[1,0]) - V_t(a,[0,1])| \leq (1-\epsilon)B.$$

Proof: Based on the third and forth equalities in Lemma 2, we only need to consider $a(t-1) = 1$. We prove by reverse induction. The inequality holds for $t = T$ since $(1-\epsilon)B$ is the maximum reward (averaged over the sensing outcome) that can be obtained in one slot. Assume that the inequality holds for $t+1$. We show that the result for t follows. Consider first $V_t(1,[1,0])$. With probability $1-\epsilon$, the users observe $K(t-1) = 1$ and choose $a(t) = 2$ according to Proposition 1. The expected immediate reward in slot t is thus $p_{0,1}(1-\epsilon)B$. With probability ϵ, we have $K(t-1) = 0$, which leads to $a(t) = 1$ and an expected immediate reward $p_{0,1}(1-\epsilon)B$ in slot t. We thus obtain $V_t(1,[1,0])$, similarly, $V_t(1,[0,1])$ as given in (10) and (11) on the next page (note that for $V_t(1,[0,1])$, we always have $K(t-1) = 0$ and $a(t) = 1$).

Applying the equalities in Lemma 2 and the upper bound on ϵ, we have

$$\begin{aligned}
&|V_t(1,[0,1]) - V_t(1,[1,0])|\\
=\ &(1-\epsilon)Bp_{0,1} - (1-\epsilon)B(\epsilon p_{1,1} + (1-\epsilon)p_{0,1})\\
&+\epsilon|V_{t+1}(1,[1,0]) - V_{t+1}(1,[0,1]|(p_{1,0}p_{0,1} - p_{1,1}p_{0,0})\\
\leq\ &2(1-\epsilon)B\epsilon(p_{0,1} - p_{1,1})\\
\leq\ &2(1-\epsilon)B\frac{p_{0,0}p_{1,1}}{p_{0,1}p_{1,0}}(p_{0,1} - p_{1,1})\\
<\ &(1-\epsilon)B,
\end{aligned}$$

where the last inequality follows from $(p_{0,1} - p_{1,1})\frac{p_{1,1}}{p_{0,1}} \leq \frac{1}{4}$ and $\frac{p_{0,0}}{p_{1,0}} < 1$.

Based on the above lemmas, we prove Theorem 1 by showing that Condition C0 in Lemma 1 holds. Let $\Omega(t) = [\omega_1(t), \omega_2(t)]$ denote the belief vector at the beginning of slot t prior to state transition. Let $\omega_i'(t)$ denote the probability that $S_i(t) = 1$, *i.e.,*

$$\omega_i'(t) = \omega_i(t)p_{1,1} + (1-\omega_i(t))p_{0,1}.$$

Assume that $\omega_1'(t) > \omega_2'(t)$, *i.e.,* the myopic action in slot t is $a(t) = 1$. We show below that $V_t^{a(t)=1}(\Omega(t)) > V_t^{a(t)=2}(\Omega(t))$.

Let $q_{i,j}$ denote the probability that $\mathbf{S}(t) = [i,j]$. Since $\omega_1'(t) > \omega_2'(t)$, we have

$$q_{1,0} = \omega_1'(t)(1-\omega_2'(t)) > q_{0,1} = (1-\omega_1'(t))\omega_2'(t). \qquad (12)$$

We now compute $V_t^{a(t)}(\Omega(t))$ by averaging over all possible values of $\mathbf{S}(t)$.

$$\begin{aligned}
V_t^{a(t)=1} &= q_{0,0}V_{t+1}(1,[0,0]) + q_{1,1}\{(1-\epsilon)B + V_{t+1}(1,[1,1])\}\\
&= q_{0,1}V_{t+1}(1,[0,1]) + q_{1,0}\{(1-\epsilon)B + V_{t+1}(1,[1,0])\}\\
V_t^{a(t)=2} &= q_{0,0}V_{t+1}(2,[0,0]) + q_{1,1}\{(1-\epsilon)B + V_{t+1}(2,[1,1])\}\\
&= q_{0,1}\{(1-\epsilon)B + V_{t+1}(2,[0,1])\} + q_{1,0}V_{t+1}(2,[1,0])
\end{aligned}$$

Applying the equalities in Lemma 2, we obtain

$$\begin{aligned}
&V_t^{a(t)=1}(\Omega(t)) - V_t^{a(t)=2}(\Omega(t))\\
=\ &(q_{1,0} - q_{0,1})\{(1-\epsilon)B + V_{t+1}(1,[1,0]) - V_{t+1}(1,[0,1])\}\\
>\ &0, \qquad (13)
\end{aligned}$$

where the last inequality follows from (12) and Lemma 3.

REFERENCES

[1] Q. Zhao and B. Sadler, "A Survey of Dynamic Spectrum Access: Signal Processing, Networking, and Regulatory Policy," *IEEE Signal Processing magazine: Special Issue on Resource-Constrained Signal Processing, Communications, and Networking*, May 2007.

[2] 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.

$$
\begin{aligned}
V_t(1,[1,0]) &= (1-\epsilon)\{p_{0,1}(1-\epsilon)B + p_{1,0}p_{0,0}V_{t+1}(2,[0,0]) + p_{1,1}p_{0,1}V_{t+1}(2,[1,1]) + p_{1,1}p_{0,0}V_{t+1}(2,[1,0]) + p_{1,0}p_{0,1}V_{t+1}(2,[0,1])\} \\
&+ \epsilon\{p_{1,1}(1-\epsilon)B + p_{1,0}p_{0,0}V_{t+1}(1,[0,0]) + p_{1,1}p_{0,1}V_{t+1}(1,[1,1]) + p_{1,1}p_{0,0}V_{t+1}(1,[1,0]) + p_{1,0}p_{0,1}V_{t+1}(1,[0,1])\} \quad (10)
\end{aligned}
$$

$$
V_t(1,[0,1]) = p_{0,1}(1-\epsilon)B + p_{0,0}p_{1,1}V_{t+1}(1,[0,0]) + p_{0,1}p_{1,1}V_{t+1}(1,[1,1]) + p_{1,1}p_{0,0}V_{t+1}(1,[0,1]) + p_{1,0}p_{0,1}V_{t+1}(1,[1,0])\} \quad (11)
$$

[3] Y. Chen, Q. Zhao, and A. Swami, "Joint Design and Separation Principle for Opportunistic Spectrum Access," in *Proc. of IEEE Asilomar Conference on Signals, Systems, and Computers*, Oct. 2006.

[4] 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.

[5] P. Papadimitratos, S. Sankaranarayanan, and A. Mishra, "A bandwidth sharing approach to improve licensed spectrum utilization," *IEEE Communications Magazine*, vol. 43, pp. 10-14, Dec. 2005.

[6] Qianchuan Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Optimal dynamic spectrum access via periodic channel sensing," to appear in *Proc. of WCNC*, 2007.

[7] Q. Zhao and A. Swami, "A decision-theoretic framework for dynamic spectrum access," in *IEEE Wireless Communications Magazine: Special Issue on Cognitive Wireless Networks*, August, 2007.

[8] S.D. Jones, N. Merheb, and I.-J. Wang, "An experiment for sensingbased opportunistic spectrum access in CSMA/CA networks," in *Proc. of the First IEEE Internatial Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, pp. 593596.

[9] S. Geirhofer, L. Tong, and B. Sadler, "Dynamic spectrum access in WLAN channels: empirical model and its stochastic analysis," in *Proc. of the First International Workshop on Technology and Policy for Accessing Spectrum (TAPAS)*, August 2006.

[10] D. Datla, A.M. Wyglinski, and G.J. Minden, "A Spectrum Surveying Framework for Dynamic Spectrum Access Networks," submitted to *IEEE Journal on Selected Areas in Communications: Cognitive Radio: Theory and Applications*, March 2007.

[11] H. L. V. Trees, "Detection, Estimation, and Modulation Theory, Part I," Wiley-Interscience, Sept., 2001.

[12] R. Smallwood and E. Sondik, "The optimal control of partially ovservable Markov processes over a finite horizon," *Operations Research*, pp. 1071–1088, 1971.