STRUCTURE AND OPTIMALITY OF MYOPIC SENSING FOR OPPORTUNISTIC SPECTRUM ACCESS

Qing Zhao

University of California
Davis, CA 95616
qzhao@ece.ucdavis.edu

Bhaskar Krishnamachari

University of Southern California
Los Angeles, CA 90089
bkrishna@usc.edu

ABSTRACT

We consider opportunistic spectrum access for secondary users over multiple channels whose occupancy by primary users is modeled as discrete-time Markov processes. Due to hardware limitations and energy constraints, a secondary user can choose, in each slot, one channel to sense and decide whether to access based on the sensing outcome. The design of sensing strategies that govern channel selections in each slot for optimal throughput performance of the secondary user can be formulated as a partially observable Markov decision process (POMDP). We exploit the structure of this problem when channels are independently and identically distributed. We reveal that the myopic sensing policy has a simple structure: channel selection is reduced to a counting process with little complexity. Further, for the two-channel case, we prove that the myopic sensing policy is in fact the optimal policy. Numerical results have also demonstrated the optimality of the myopic sensing policy when there are more than two channels.

Index Terms: Opportunistic spectrum access, POMDP, myopic policy.

1. INTRODUCTION

Opportunistic spectrum access (OSA), where secondary users identify and exploit local and instantaneous spectrum availability while limiting interference to primary users (licensees), is one of the approaches envisioned for dynamic spectrum management [1, 2]. A basic component of OSA is a sensing strategy at the MAC layer. Due to hardware limitations and energy constraints, a secondary user may not be able to sense all channels in the spectrum simultaneously. In this case a sensing strategy for intelligent channel selection is necessary to track the rapidly varying spectrum opportunities.

The purpose of a sensing strategy is twofold: catch a spectrum opportunity for immediate access and obtain statistical information on spectrum occupancy so that more rewarding sensing decisions can be made in the future. A tradeoff has to be reached between these two often conflicting objectives, and the design of optimal sensing strategies is, in general, a sequential decision making problem.

By modeling primary users' channel occupancy as Markov processes, the design of sensing strategies can be formulated as a partially observable Markov decision process (POMDP). Unfortunately, obtaining the optimal policy for a general POMDP is often intractable. For the OSA problem, the complexity of obtaining the optimal sensing policy can be shown to be $\mathcal{O}(N^T)$, where N is the number of channels in the spectrum of interest and T is the time horizon length [3].

A common approach of trading performance for tractable solutions is to consider myopic policies. A myopic policy aims solely at maximizing the immediate reward, ignoring the impact of the current action on the future reward. Obtaining myopic policies is thus a static optimization problem instead of a sequential decision making problem. As a consequence, the complexity is significantly reduced, often at the price of considerable performance loss.

In this paper, we show that for designing sensing strategies in OSA, low complexity does not necessarily imply suboptimal performance. The myopic sensing policy with a simple structure achieves the optimal performance when channels are independently and identically distributed (i.i.d.).

The contribution of this paper is twofold. First, we reveal a simple structure of the myopic sensing policy for i.i.d. channels. Specifically, selecting channels in each slot is reduced to a simple counting procedure: a secondary user only needs to set up pointers indicating the channels to which the last visits occurred most recently or the longest time ago. Second, we prove that for the two-channel case, this myopic sensing policy with such a simple structure is actually optimal. The optimality of the myopic sensing policy in cases with more than two channels has also been demonstrated via extensive numerical examples.

Related Work The majority of existing work on OSA focuses on the spatial domain where spectrum opportunities are considered static or slowly varying in time. As a con-

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.

sequence, real-time opportunity identification and tracking is not as critical a component in this class of applications, and the prevailing approach tackles network design in two separate steps: (i) opportunity identification assuming continuous full-spectrum sensing; (ii) opportunity allocation among secondary users assuming perfect knowledge of spectrum opportunities at any location over the entire spectrum. Opportunity identification in the presence of fading and noise uncertainty has been studied in [4,5]. Spatial opportunity allocation among secondary users can be found in [6–8] and references therein.

Exploiting temporal spectrum opportunities that also vary in space requires a joint design of spectrum detectors at the physical layer and spectrum sensing and access strategies at the MAC layer [9]. Tracking the rapidly varying spectrum opportunities becomes a critical issue, which is the focus of this paper. Clearly, a simple yet sufficiently accurate statistical model of spectrum occupancy is crucial to the efficiency of spectrum opportunity tracking. Measurements obtained from spectrum monitoring test-beds [10] demonstrate the Markovian transition between busy and idle channel states in 802.11b, a similar model as used in this paper.

The POMDP framework for OSA was first proposed in [11]. The goal of this paper is to investigate the structure and optimality of myopic sensing. An overview of challenges and recent developments in OSA can be found in [12].

2. 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. The traffic statistics of the primary network are such that the occupancy of these channels follows N independent discrete-time Markov processes with 2 states. Specifically, the state of channel i in slot t is given by

$$S_i(t) \in \{0 \text{ (occupied)}, 1 \text{ (idle)}\}.$$

The state diagram and transition probabilities $\{p_{i,j}^{(n)}\}$ of channel n are illustrated in Figure 1. We assume that the spectrum usage statistics of the primary network remain unchanged for T slots.

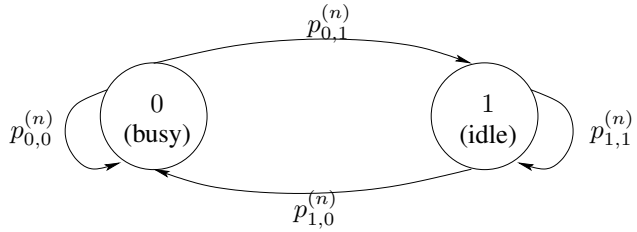

Fig. 1. The Markov channel model

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. If the channel is sensed to be idle, the user transmits using carrier sensing. We assume here that sensing errors are negligible. A sensing strategy that dynamically selects channels for intelligent spectrum opportunity tracking is crucial to the throughput performance of the secondary user. The focus of this paper is the design of spectrum sensing strategies that achieve a favorable tradeoff between performance and complexity. Details of the protocol implementation can be found in [13], and have been omitted in this paper due to space limit.

We assume that the transition probabilities of the Markovian model have been learned by the secondary user. Results on the robustness to model mismatch can be found in [9]. If the transition probabilities are unknown, formulations and algorithms for POMDP with an unknown model exist in the literature [14] and can be applied to this problem.

3. A POMDP FORMULATION

In this section, we present the POMDP formulation of designing spectrum sensing strategies.

3.1. Reward and Design Objective

Let π_s denote a sensing policy that decides, sequentially, which channel to sense in each slot. The design of π_s can be formulated as a POMDP as first given in [11]. Specifically, the underlying state space is $\mathbf{S} = [S_1, \cdots, S_N] \in \{0,1\}^N$, and the action space is $a \in \{1, \cdots, N\}$. A natural choice of reward $R_a(t)$ for choosing channel a in slot t is[1]

$$R_a(t) = S_a(t)B_a,$$

which represents the number of bits delivered. We can then define the objective function as the total number of bits transmitted in T slots:

$$J \triangleq \mathbb{E}_{\pi_s}[\sum_{t=1}^{T} R_a(t)], \tag{1}$$

where $\mathbb{E}_{\pi_s}$ represents the conditional expectation given that a sensing policy π_s is employed. Note that the reward $R_a(t)$ obtained in slot t depends on the sensing action (which channel a to sense) and the state of the underlying Markov process (channel availability) in slot t.

3.2. Sufficient Statistics

Since the system state $\mathbf{S}$ cannot be directly observed, we can only infer it from partial sensing outcomes. It has been show in [15] that the *a posterior* distribution of the system state

[1] In this paper, we focus on distributed sensing strategies where secondary users make independent and selfish decisions without coordination. In this case, a secondary user chooses its spectrum sensing strategies under the assumption that it will receive a reward when the chosen channel is not used by the primary network.

that exploits the entire sensing and decision history characterizes our knowledge about the system state and is a sufficient statistic for optimal decision making. Specifically, at the beginning of slot t, our knowledge of the system state based on all past decisions and observations can be summarized by a belief vector

$$\Lambda(t) = [\lambda_1(t), \cdots, \lambda_{2^N}(t)]$$

where $\lambda_j(t)$ is the conditional probability (given the decision and observation history) that the system state is j ($j = 1, \cdots, 2^N$) at the beginning of slot t. Note that $\Lambda(t)$ characterizes the system state in slot t *prior* to the state transition (a conventional notation in the literature of POMDP).

For independently evolving channels, it has been shown in [11, 13] that the marginal conditional distribution is a sufficient statistic, *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.

3.3. Optimal Sensing Policy

With the concept of belief vector, a sensing policy π_s essentially defines the mapping from $\Omega(t)$ to the index a of the channel to be sensed for each slot t:

$$\begin{aligned} \pi_s &= [\mu_1, \cdots, \mu_T], \\ \text{where } \mu_t &: \Omega(t) \in [0,1]^N \rightarrow a \in \{1, \cdots, N\}. \end{aligned}$$

The design objective is to choose π_s to maximize the throughput J. Note that for a POMDP over a finite horizon, the optimal policy is generally non-stationary; the mapping from the belief vector to the optimal action varies over time. Clearly, the complexity of obtaining the optimal (non-stationary) policy grows with the horizon length T.

Referred to as the value function, $V_t(\Omega(t))$ denotes the maximum expected remaining reward that can be accrued starting from slot t when the current belief vector is $\Omega(t)$. It has two parts: (i) the immediate reward $R_a(t)$ obtained in slot t when the user senses channel a; (ii) the maximum expected remaining reward $V_{t+1}(\Omega(t+1))$ starting from slot $t+1$ given a belief vector

$$\Omega(t+1) = \mathcal{T}(\Omega(t)|a, S_a(t))$$

which represents the updated knowledge of the system state after incorporating the action a and the observed channel state $S_a(t)$ in slot t. Averaging over all possible system states and observations, we arrive at the following Bellman's equation

$$V_t(\Omega(t)) = \max_{a=1,\cdots,N} \mathbb{E}[R_a(t) + V_{t+1}(\mathcal{T}(\Omega(t)|a, S_a(t)))], \quad (2)$$

where the updated belief vector $\Omega(t+1) = \mathcal{T}(\Omega(t)|a, S_a(t))$ can be easily obtained via the Bayes rule.

From (2) we can see that an action chosen at a slot affects the total reward in two ways: it acquires an immediate reward $R_a(t) = S_a(t)B$ in this slot and transforms the belief vector to $\mathcal{T}(\Omega|a, S_a(t))$ which determines the future reward $V_{t+1}(\mathcal{T}(\Omega(t)|a, S_a(t)))$. The optimal policy strikes a balance between gaining immediate reward and gaining information for future use. However, due the impact of the current action on the future reward, the uncountable belief space, and the non-stationary nature of the optimal policy, obtaining the optimal solution to a general POMDP is often computationally prohibitive. For the OSA problem, the complexity of obtaining the optimal sensing policy is $\mathcal{O}(N^T)$, which grows exponentially with T [3].

4. THE MYOPIC POLICY

For tractable solutions, one often resorts to a myopic policy that ignores the impact of the current action on the future reward, focusing solely at maximizing the immediate reward. Myopic policies are thus stationary; the mapping from the belief vector to the myopic action is the same for all slots. For the spectrum access problem at hand, the action chosen by the myopic sensing policy is given by [11]

$$a_*(t) = \arg \max_{a=1,\cdots,N} (\omega_a(t)p_{1,1}^{(a)} + (1-\omega_a(t))p_{0,1}^{(a)})B_a, \quad (3)$$

where $(\omega_a(t)p_{1,1}^{(a)} + (1-\omega_a(t))p_{0,1}^{(a)})$ is the probability that channel a is available in slot t (note that $\omega_a(t)$ denotes the channel availability probability prior to state transition).

At the end of slot t, the belief vector Ω is updated based on the action $a_*(t)$ and the observed channel state $S_{a_*}(t)$ as in (4) shown on the next page. Note that when a channel is not sensed, the probability of its availability is updated according to the Markov chain. If the channel is sensed, the state of this channel is the sensing outcome.

We show in the next section that for i.i.d. channels, we do not need to update the belief vector in each slot as given in (4). Obtaining the optimal myopic action $a_*(t)$ is reduced to a simple counting process.

5. THE STRUCTURE AND OPTIMALITY OF MYOPIC SENSING FOR I.I.D. CHANNELS

In this section, we show that when we have N identical channels ($p_{i,j}^{(n)} \equiv p_{i,j},\ B_n \equiv B$), the myopic sensing policy has an interesting structure that leads to further complexity reduction. We further prove that the myopic sensing policy is optimal when $N = 2$, *i.e.,* the optimal performance can be achieved at little complexity due to the strong structure and the optimality of the myopic policy in this case. For $N > 2$, extensive numerical results have demonstrated the optimality of the myopic policy. We are currently extending the opti-

$$\omega_i(t+1) = \begin{cases} 1 & \text{if } a_*(t) = i, S_{a_*}(t) = 1 \\ 0 & \text{if } a_*(t) = i, S_{a_*}(t) = 0 \\ \omega_i(t)p_{1,1}^{(a)} + (1-\omega_i(t))p_{0,1}^{(a)} & \text{if } a_*(t) \neq i \end{cases}. \quad (4)$$

mality proof for $N = 2$ to general cases with the aid of the structure of the myopic policy revealed in Theorem 1 below.

5.1. The Structure of Myopic Sensing

We give in Theorem 1 the structure of the myopic sensing policy for i.i.d. channels. The structure is given for $p_{0,1} > p_{1,1}$, *i.e.,* a channel is more likely to change from busy (0) to idle (1) than staying idle. For the case where $p_{0,1} < p_{1,1}$, we can simply switch these two states and replace $p_{0,1}$ and $p_{1,1}$ by $1 - p_{0,1}$ and $1 - p_{1,1}$, respectively. We then go back to the former case. When $p_{0,1} = p_{1,1}$, the dynamics of channel states degenerate from a Markov process to an i.i.d sequence. In this case, no information on the current channel state can be obtained from the sensing and decision history. The optimal sensing policy is to simply choose any channel in each slot.

Theorem 1 *Consider N i.i.d. channels with $p_{0,1} > p_{1,1}$. In slot t, let $\delta_i(t) \in \{1, 2, \cdots, t-1, \infty\}$ denote the time difference between t and the last visit to channel i. If channel i has never been visited, then $\delta_i(t) = \infty$. Define the following sets.*

$$\begin{aligned} \Delta_e(t) &\triangleq \{\tau_i(t) : \tau_i(t) \text{ is even}\}, \\ \bar{\Delta}_e(t) &\triangleq \{\tau_i(t) : \tau_i(t) \text{ is odd or } \infty\}. \end{aligned}$$

Given the action $a(t-1)$ and sensing outcome $S_{a(t-1)}(t-1)$ in slot $t-1$, the myopic action $a_*(t)$ in slot t that maximizes the expected immediate reward (see (3)*) is as follows.*

$$a_*(t) = \begin{cases} a(t-1) & \text{if } S_{a(t-1)}(t-1) = 0 \\ \arg\min \Delta_e(t) & \text{if } S_{a(t-1)}(t-1) = 1, \Delta_e(t) \neq \emptyset \\ \arg\max \bar{\Delta}_e(t) & \text{if } S_{a(t-1)}(t-1) = 1, \Delta_e(t) = \emptyset \end{cases}. \quad (5)$$

The proof of Theorem 1 is based on the eigen-decomposition of the k-step transition matrix P^k. Details are given in Appendix I.

Theorem 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 when a 0 is observed in the previous slot and switch to another channel when a 1 is observed. 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 special case of $N = 2$, the myopic policy is simply to stay in the same channel after observing 0 and switch to the other channel after observing 1.

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 Theorem 1 when the initial distribution is not the stationary distribution.

Theorem 1 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.

A natural question that follows is how much performance has to be sacrificed in order to use a sensing strategy with such a simple structure, which is addressed next.

5.2. The Optimality of Myopic Sensing

Surprisingly, extensive numerical results have demonstrated that the myopic sensing policy achieves the optimal performance for i.i.d. channels. One example is given in Figure 2, where we compare the throughput performance of the myopic sensing and the optimal policy [13]. We consider 3 independent channels. In the upper figure, these channels are identical, while in the lower figures, channels have different transition matrixes. 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.

Exploiting the structure of myopic sensing given in Theorem 1, we have proven its optimality for $N = 2$ i.i.d. channels as given in Theorem 2. We are currently extending the proof to general cases with $N > 2$.

Theorem 2 *For two i.i.d. channels, the myopic sensing policy is optimal.*

The proof of Theorem 2 is based on the following lemma which applies to any POMDP over a finite horizon. Details are given in Appendix II.

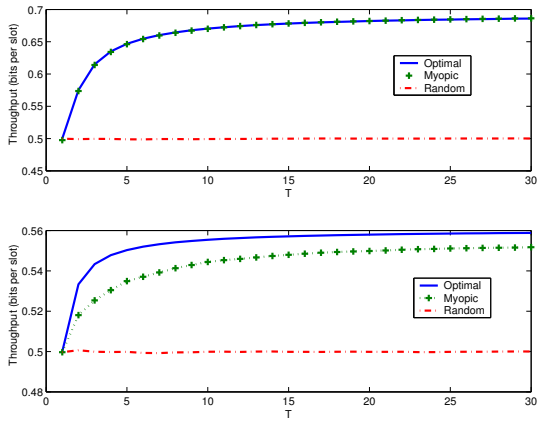

Fig. 2. Performance of myopic sensing.

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 $t+1, \cdots, T$.

6. CONCLUSION AND FUTURE WORK

In this paper, we address spectrum opportunity tracking within the framework of POMDP. In particular, we demonstrate the strong structure and optimal performance of the myopic sensing strategy when channels are i.i.d. On-going work includes the extension of the optimality proof for the two-channel case to general cases with more than two channels and in the presence of sensing errors. Quantitatively characterizing the performance loss of the myopic sensing strategy for non-i.i.d. channels is also among the future research directions.

ACKNOWLEDGEMENT

The authors wish to thank Yunxia Chen at UC Davis for her discovery of the optimality of myopic sensing in i.i.d. channels based on extensive numerical examples.

APPENDIX I: PROOF OF THEOREM 1

Consider first $N = 2$. Without loss of generality, assume $a(t-1) = 1$. Consider first $S_1(t-1) = 0$. The immediate reward for staying in channel 1 in slot t is $p_{0,1}B$, while the immediate reward for switching to channel 2 in slot t is

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

where the inequality follows from $p_{0,1} > p_{1,1}$. Hence, the myopic action in slot t is to stay in channel 1 when $S_1(t-1) = 0$.

Similarly, when $S_1(t-1) = 1$. 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

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

Hence, the myopic action in slot t is to switch to channel 2 when observing 1.

Applying the above arguments to $N > 2$, we know that the myopic action is to stay after observing 0 and switch after observing 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 1, the last known state of every channel is 1. Hence, if $S_{a(t-1)}(t-1) = 1$, we should switch to channel a_* given by

$$a_* = \arg\max_i p_{1,1}^{\delta_i(t)},$$

where $p_{1,1}^{\delta_i(t)}$ denotes the probability of staying in state 1 after $\delta_i(t)$ slots. Based on the eigen-decomposition of the k-step transition matrix, we obtain the following expression of $p_{1,1}^k$ [16].

$$p_{1,1}^k = \frac{p_{0,1} + p_{1,0}(p_{1,1} - p_{0,1})^k}{p_{0,1} + p_{1,0}}.$$

For $p_{1,1} < p_{0,1}$, it is easy to see that $p_{1,1}^k$ is decreasing to the stationary distribution for even k's and increasing to the stationary distribution for odd k's. Theorem 1 thus follows.

APPENDIX II: PROOF OF THEOREM 2

We first prove Lemma 1 by reverse induction. We will show that the optimal action in every slot is myopic. For slot T, the optimal action is clearly myopic. Assume that for slots $t+1, \cdots, T$, the optimal actions are myopic. Based on Condition C0, the optimal action in slot t is also myopic. Lemma 1 thus follows by induction.

Next, we prove Theorem 2 by showing that Condition C0 holds. Let $V_{t+1}^{a(t)}(\mathbf{S}(t))$ denote the total expected remaining reward starting at slot $t+1$ under myopic actions for a given $\mathbf{S}(t)$. We will show that $V_{t+1}^{a(t)}(\mathbf{S}(t))$ is completely determined by $\mathbf{S}(t)$, *i.e.,* independent of the action $a(t)$ in slot t, and

$$V_{t+1}^{a(t)}([0,1]) = V_{t+1}^{a(t)}([1,0]).$$

This is a stronger condition than C0.

Prove by induction. We show first the condition holds for slot $T-1$ and $V_T(0,1) = V_T(1,0)$. Consider first $\mathbf{S}(T-1) = [0,0]$ or $[1,1]$. Due to symmetry in the state and in the dynamics of the two Markov chains, the two possible actions $a(T-1) = 1$ and $a(T-1) = 2$ are indistinguishable; they lead to the same maximum expected reward in slot T. For $\mathbf{S}(T-1) = [0,1]$, with $a(T-1) = 1$, the myopic action in slot T is to stay in channel 1 (see Theorem 1). The resulting expected reward in slot T is $p_{0,1}B$. With $a(T-1) = 2$, the

myopic action in slot T is to switch to channel 1, resulting in the same expected reward $p_{0,1}B$. The same line of arguments applies to $\mathbf{S}(T-1) = [1,0]$, and the maximum expected reward in slot T is also $p_{0,1}B$, *i.e.,* $V_T(0,1) = V_T(1,0)$.

Assume that the claim holds for slot t where $t < T-1$, *i.e.,* $V_{t+1}^{a(t)}(\mathbf{S}(t))$ is independent of $a(t)$ and $V_{t+1}([0,1]) = V_{t+1}([1,0])$. We show next the condition holds for slot $t-1$. Consider first $\mathbf{S}(t-1) = [0,0]$. We obtain $V_t(0,0)$ under each action by considering all 4 possible states in slot t.

$$\begin{aligned} V_t^{a=1}(0,0) &= p_{0,0}^2(0 + V_{t+1}(0,0)) + p_{0,1}^2(B + V_{t+1}(1,1)) \\ &+ p_{0,0}p_{0,1}(0 + V_{t+1}(0,1)) + p_{0,1}p_{0,0}(B + V_{t+1}(1,0)) \\ V_t^{a=2}(0,0) &= p_{0,0}^2(0 + V_{t+1}(0,0)) + p_{0,1}^2(B + V_{t+1}(1,1)) \\ &+ p_{0,0}p_{0,1}(B + V_{t+1}(0,1)) +_{0,1} p_{0,0}(0 + V_{t+1}(1,0)) \end{aligned}$$

Since $V_{t+1}(0,0)$, $V_{t+1}(0,1)$, $V_{t+1}(1,0)$, and $V_{t+1}(1,1)$ are independent of actions taken in slot t and $V_{t+1}(0,1) = V_{t+1}(1,0)$, we see that $V_t^{a=1}(0,0) = V_t^{a=2}(0,0)$. Similarly, we reach the same statement for $V_t(0,1)$, $V_t(1,0)$, and $V_t(1,1)$. By comparing $V_t(0,1)$ and $V_t(1,0)$, we also obtain $V_t(0,1) = V_t(1,0)$.

7. REFERENCES

[1] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Proc. IEEE International Workshop on Mobile Multimedia Communications*, pp. 3–10, 1999.

[2] "DARPA: The Next Generation (XG) Program." http://www.darpa.mil/ato/programs/xg/index.htm.

[3] D. Djonin, Q. Zhao, and V. Krishnamurthy, "Optimality and Complexity of Opportunistic Spectrum Access: A Truncated Markov Decision Process Formulation," in *Proc. of International Conference on Communications (ICC)*, June 2007.

[4] A. Sahai and N. Hoven and R. Tandra, "Some fundamental limits on cognitive radio," in *Proc. Allerton Conference on Communication, Control, and Computing*, October 2004.

[5] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proc. the 38th. Asilomar Conference on Signals, Systems, and Computers*, pp. 772 – 776, 2004.

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

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

[8] S. Sankaranarayanan, P. Papadimitratos, A. Mishra, and S. Hershey, "A Bandwidth Sharing Approach to Improve Licensed Spectrum Utilization," in *Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2005.

[9] 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.

[10] 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.

[11] Q. Zhao, L. Tong, and A. Swami, "Decentralized cognitive MAC for dynamic spectrum access," in *Proc. of IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Nov. 2005.

[12] 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.

[13] 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.

[14] D. Aberdeen, "A survey of approximate methods for solving partially observable markov decision processe," tech. rep., National ICT Australia, December 2003. `http://users.rsise.anu.edu.au/~daa/papers.html`.

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

[16] R. G. Gallager, *Discrete Stochastic Processes*. Kluwer Academic Publishers, 1995.