Optimality and Complexity of Opportunistic Spectrum Access: A Truncated Markov Decision Process Formulation

Dejan V. Djonin, Qing Zhao, *IEEE, Member*, Vikram Krishnamurthy, *IEEE, Fellow*

***Abstract*— We consider opportunistic spectrum access (OSA) which allows secondary users to identify and exploit instantaneous spectrum opportunities resulting from the bursty traffic of primary users. Within the framework of Partially Observable Markov Decision Process (POMDP), we develop decentralized cognitive MAC protocols that allow secondary users to independently search for spectrum opportunities without a central coordinator or a dedicated communication channel. The focus of this paper is the tradeoff between optimality and complexity of obtaining OSA protocols. We first analyze the computational complexity of designing OSA protocols within the POMDP framework and demonstrate that the complexity grows exponentially with the horizon length (*ie,* the spectrum access time of secondary users). By exploiting the underlying structure of the problem, we aim to develop a quantitative characterization of the fundamental tradeoff between optimality and complexity so that a systematic way of balancing these two can be obtained. Specifically, by exploiting the mixing time of the underlying Markov process of spectrum occupancy, we develop a truncated MDP formulation of OSA and reduce the computational complexity from growing exponentially to linearly with the horizon length. More importantly, this truncated MDP formulation provides a systematical way of trading off performance with complexity by choosing an appropriate truncation parameter.**

Keywords: Opportunistic Spectrum Access, Optimal Control, Markov Decision Processes, Imperfect State Information, History Truncation

I. INTRODUCTION

A. Opportunistic Spectrum Access and Prevailing Approach

While the current state of spectrum allocation indicates that almost all usable radio frequencies have already been occupied, extensive measurements reveal the pervasive existence of spectrum opportunities in both time and space. It has been shown that, at any given time and location, a large portion of licensed spectrum lies unused. For example, over 62% white space exists in the spectrum under 3GHz [1]. Even when a frequency band is actively used, the bursty arrivals of many applications result in abundant spectrum opportunities at the millisecond scale. Traffic measurements of wireless LAN taken at Cornell University indicate 75% idle time during an active FTP session. For voice-over-IP applications such as Skype, up to 90% idle time has been observed.

These measurements of actual spectrum usage have highlighted the drawbacks of the current static spectrum allotment policy. They also form the key rationale for Opportunistic Spectrum Access (OSA) envisioned by the DARPA XG program [2] and also considered by FCC [3], [4]. The idea is to exploit instantaneous spectrum availability by opening licensed spectrum to secondary users. This would allow secondary users to identify available spectrum resources and communicate in a manner that limits the level of interference perceived by primary users.

The prevailing approach to OSA tackles network design in two separate steps: (i) opportunity identification assuming continuous full-spectrum sensing; (ii) opportunity allocation among secondary users assuming full knowledge of spectrum opportunities. Opportunity identification in the presence of fading and noise uncertainty has been studied in [5], [6]. Decentralized opportunity allocation strategies can be found in [7]–[9] and references therein.

B. Scope and Contribution

In this paper, we address the exploitation of temporal spectrum opportunities resulting from the bursty traffic of primary users. Given the rapid temporal variations of spectrum occupancy, opportunity identification and opportunity exploitation need to be designed jointly so that secondary users can make real-time decisions on spectrum access.

We consider an ad hoc OSA network where there is no central coordinator or a dedicated communication/control channel. Recognizing hardware and energy constraints of low-cost battery-powered wireless nodes, we assume that a secondary user may not be able to perform full-spectrum sensing or may not be willing to monitor the spectrum when it has no data to transmit.

To exploit temporal opportunities in a spectrum only part of which can be sensed at any given time, secondary users must make real-time decisions on where in the spectrum to sense and whether access is possible based on the sensing outcome. As shown in [10], [11], such decisions hinge on the optimal use of statistical information about the dynamics of spectrum occupancy obtained from previous sensing outcomes. Adopting a Markovian model of spectrum occupancy, the authors of [10], [11] have developed an analytical framework based on the theory of Partially Observable Markov Decision Process (POMDP) that integrates spectrum sensing with spectrum access.

The complexity of POMDP mainly comes from the continuously growing observation history. To achieve a favorable tradeoff between optimality and complexity, we develop methods of truncating the observation history without decimating the performance. The key is to exploit the mixing time of the underlying Markov process of spectrum occupancy[1]. By truncating the observation history according to the mixing time, we develop a truncated MDP formulation of OSA and reduce

---

[0] This work was supported in part by the NSERC PostDoctoral Fellowship Award, the NSERC strategic grant, by 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.
D.V. Djonin and V. Krishnamurthy are with the Department of Electrical and Computer Engineering, 2356 Main Mall, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada, Phone: +1 604 827 3268, +1 604 822 2653, Fax: +1 604 822 5949, E-mails: ddjonin@ece.ubc.ca; vikramk@ece.ubc.ca. Q. Zhao is with the Department of Electrical and Computer Engineering, University of California, Davis, CA. E-mail: qzhao@ece.udavis.edu

[1] The mixing time of a finite-state irreducible and aperiodic Markov process characterizes how fast the process converges to its stationary distribution. The convergence is geometrical with rate determined by the largest absolute value of the non-trivial eigenvalues of the transition matrix.

the computational complexity from growing exponentially to linearly with the horizon length. More importantly, this truncated MDP formulation provides a systematical way of trading off performance with complexity by choosing an appropriate truncation parameter.

The rest of the paper is organized as follows. In Section II, we present the network model. In Section III, we analyze the computational complexity of obtaining the optimal OSA protocol within the POMDP framework proposed in [10], [11]. We then develop the truncated MDP formulation in Section IV and analyze its complexity. Extensions and discussions are given in Section V, followed by numerical examples in Section VI. The paper is concluded in Section VII.

II. PROBLEM STATEMENT

A. *The Network Model*

Consider a spectrum consisting of N channels[2], each with bandwidth B_i $(i = 1, \cdots, N)$. These N channels are allocated 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 N channels follows a discrete-time Markov process with 2^N states. Specifically, the network state in slot t is given by $[X_1(t), \cdots, X_N(t)]^t$ where $X_i(t) \in \{0 \text{ (occupied) }, 1 \text{ (idle) }\}$. A sample path of the state evolution for $N = 2$ is illustrated in Figure 1.

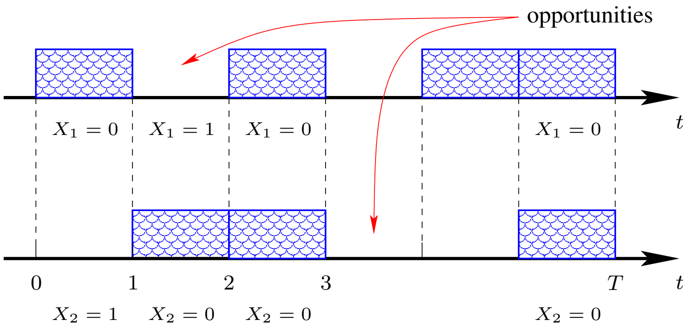

Fig. 1. A sample path of spectrum occupancy.

We assume that the spectrum usage statistics of the primary network remain unchanged for T slots. We also assume in this paper that the transition probabilities of the Markovian model are known. A study of the impact of mismatched model on the performance of OSA can be found in [12]. When the transition probabilities are unknown, protocol design for OSA can be formulated as a POMDP with unknown model and existing results in the literature can be applied (see [14] and references therein).

We consider a secondary network seeking spectrum opportunities in these N channels (see Figure 1). We focus on an ad hoc network where secondary users join/exit the network and sense/access the spectrum independently without exchanging local information. In each slot, a secondary user chooses a set of channels to sense and a set of channels to access. Limited by its hardware constraints and energy supply, a secondary user can sense no more than L_1 $(L_1 \leq N)$ and access no more than L_2 $(L_2 \leq L_1)$ channels in each slot.

III. A POMDP FORMULATION FOR OPTIMALITY

In this section, we present the POMDP formulation of OSA first developed in [10]. We then analyze the complexity of obtaining the optimal OSA protocols within the POMDP framework.

A. *A POMDP Formulation*

To simply notation, we consider the case where a secondary user can sense and access one channel in each slot $(L_1 = L_2 = 1)$. The formulations and results obtained in this paper, however, apply to general cases as discussed in Section V. We also assume that sensing errors are negligible. In this case, a secondary user transmits if and only if the channel is sensed to be available. Collisions with primary users are thus avoided. In the presence of sensing errors, secondary users need to design carefully the access strategy by taking into account the operating characteristics (false alarm vs. miss detection) of the spectrum sensor. The optimal joint design of the spectrum sensor, the sensing strategy for channel selection, and the access strategy based on imperfect sensing outcomes is addressed in [12], [13].

The problem of designing the optimal OSA protocol is the problem of computing the optimal control of a partially observed (hidden) Markov chain and is commonly referred to as a Partially Observed Markov Decision Process (POMDP). Specifically, this POMDP consists of

- Decision intervals: time slots $\{1, \ldots, T\}$;
- States $\mathcal{S} = \{1, \ldots, 2^N\}$ denoting the availability of each channel;
- Transition probabilities of the underlying Markovian model of spectrum occupancy: $p_{i,j}, i, j \in \mathcal{S}$;
- Action $a \in \{1, \ldots, N\}$: sense channel a and access it if available;
- Observation $y_{s,a} \in \{0, 1\}$ denoting the availability of the sensed channel a for state $s \in \mathcal{S}$;
- Reward $c(a, y)$ of choosing action a and receiving observation y.

At the beginning of a decision interval, the state of each channel evolves according to the transition probability matrix $\{p_{i,j}\}$. For a chosen action a which specifies the channel to be sensed in this decision interval, a secondary user senses the channel and transmits if the chosen channel is available. The outcome of channel sensing is the observation $y_{s,a}$ which indicates the availability of the chosen channel. The additive reward $c(a, y)$ that depends on the action a and observation y is accrued at the end of this decision interval. In the context of opportunistic spectrum access, $c(a, 0) = 0$ and $c(a, 1) = B_a$, indicating the number of bits transmitted by the user when the channel is available (which is proportional to the bandwidth B_a of the channel)[3].

The objective is to choose the sensing action a sequentially in each slot so that the throughput of the secondary user over T slots is maximized. Namely, the knowledge of the internal state of the system based on all past decisions and observations can be encoded as an information vector $\pi = [\pi_1, \ldots, \pi_{2^N}]$ where π_i is the conditional probability (based on previous decisions and observations) that the state of the system is i at the beginning of the current decision slot. It has been shown in [15]

---

[2] Here we use the term channel broadly. A channel can be a frequency band with certain bandwidth. It can also be a collection of spreading codes in a CDMA network or a set of tones in an OFDM system.

[3] We focus on decentralized cognitive MAC where secondary users make independent and selfish decisions without coordination. In this case, a secondary user chooses its sensing action under the assumption that it will receive a reward when the chosen channel is not used by the primary network.

that at any time the information vector π is a sufficient statistic for the optimal policy. Furthermore, the statistical behavior of the information state π is that of a discrete-time continuous-state Markov process.

A control policy μ is thus a sequence of functions, each mapping the current information state $\pi(t)$ to the action $a \in \{1, 2, \ldots, N\}$ to be taken in slot t.

$$\begin{aligned} &\mu = [\mu_1, \cdots, \mu_T], \\ &\text{where } \mu_t : \pi(t) \in [0,1]^{2^N} \longrightarrow a \in \{1, \ldots, N\}. \end{aligned} \tag{1}$$

The optimal control μ^*, which defines the optimal OSA protocol, is the one that maximizes the expected reward over the horizon of length T conditioned on the initial information state π_0, *ie,*

$$\begin{aligned} J_\mu &= \mathbb{E}\left[\sum_{t=1}^{T} c(\mu_t(\pi(t)), y(t)) | \pi_0\right], \\ \mu^* &= \arg\max J(\mu). \end{aligned} \tag{2}$$

B. Complexity Analysis

We now provide a complexity analysis of obtaining the optimal OSA protocol under the POMDP formulation.

Referred to as the value function, $V_t(\pi(t))$ denotes the maximum expected remaining reward that can be accrued starting from slot t when the current information vector is $\pi(t)$. It has two parts: (i) the immediate reward $c(a, y)$ obtained in slot t; (ii) the maximum expected remaining reward $V_{t+1}(\pi(t+1))$ starting from slot $t+1$ given an information vector $\pi(t+1) = \mathcal{T}(\pi(t)|a, y)$ which represents the updated knowledge of the system state after incorporating the action a and observation y obtained in slot t. Averaging over all possible system states and observations, we arrive at the following Bellman's equation

$$\begin{aligned} V_t(\pi(t)) \;=\; &\max_{a=1,\cdots,N} \{\sum_{i=1}^{2^N} \pi_i(t) \sum_{j=1}^{2^N} p_{i,j} \sum_{\theta=0}^{1} \Pr[y_{j,a} = \theta] \\ &(c(a, \theta) + V_{t+1}(\mathcal{T}(\pi(t)|a, \theta)))\}, \end{aligned} \tag{3}$$

where $\pi_i(t)$ denotes the ith entry of $\pi(t)$, and the updated information vector $\pi(t+1) = \mathcal{T}(\pi(t)|a, y)$ can be easily obtained via the Bayes rule.

From (3) we can see that an action chosen at a slot affects the total reward in two ways: it acquires an immediate reward $c(a, y)$ in this slot and transforms the belief vector to $\mathcal{T}(\pi|a, y)$ which determines the future reward $V_{t+1}(\mathcal{T}(\pi(t)|a, y))$. The optimal policy strikes a balance between gaining instantaneous reward and gaining information for future use.

The optimal policy of the POMDP can be computed in an iterative fashion based on (3). The computational complexity is determined by the number of possible values of the information vector $\pi(t)$ for each t, which is analyzed below.

Let $|S_t|$ denote the number of possible values of the information state vector in slot t. The number of actions is given by the number N of channels. Under each action, one information state vector can transit to two information state vectors according to the observation 0 or 1. The complexity in slot t is thus given by $2N|S_t|$. Notice that in the first slot, the information state vector is the stationary distribution, and in each slot, one information state vector can transit to $2N$ possible information state vectors (for N actions and 2 observations under each action), we have

$$|S_1| = 1, |S_t| = 2N|S_{t-1}| = (2N)^{t-1}. \tag{4}$$

Since the complexity in the last slot is N, the total complexity of obtaining the optimal POMDP policy over a horizon of length T is given by

$$\sum_{t=1}^{T-1} 2N|S_t| + N = \sum_{t=1}^{T-1} (2N)^{t-1} + N \approx (2N)^{T-1}, \tag{5}$$

which grows exponentially with the horizon length T.

Note that in this complexity analysis, we consider only a fixed initial information vector π. The complexity of obtaining the complete optimal policy (consider all distribution vectors) would be on the order of N^{2^T}.

IV. A TRUNCATED MDP FORMULATION FOR FAVORABLE TRADEOFF BETWEEN OPTIMALITY AND COMPLEXITY

In this section, we develop a truncated MDP formulation of OSA by exploiting the mixing time of the underlying Markov process of spectrum occupancy. This formulation allows us to reduce the computational complexity of the OSA protocol from growing exponentially to linearly with the horizon length T. More importantly, this truncated MDP formulation provides a systematical way of trading off performance with complexity by choosing an appropriate truncation parameter.

A. A Discrete Sufficient Statistic with Reduced Dimension

We consider first N independently evolving channels. Extensions to dependent channels are discussed in Section V. The transition probabilities of channel i are given by $\{\alpha_i, \beta_i\}, i = 1, \ldots, N$. A transition diagram of channel i is shown in Figure 2. The transition matrix P_i of channel i is given by

$$P_i = \left[\begin{array}{cc} 1-\alpha_i & \alpha_i \\ \beta_i & 1-\beta_i \end{array}\right] \tag{6}$$

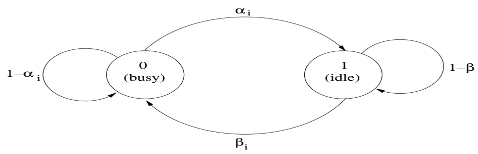

Fig. 2. Markovian channel model for the i-th channel.

The first step to complexity reduction is to develop a sufficient statistic with a dimension linear with respect to N instead of exponential (2^N) as the information vector π. For N independent channels, it has been shown in [10] that a sufficient statistic is given by $\tilde{\pi} = [\tilde{\pi}_1, \ldots, \tilde{\pi}_N]$, where $\tilde{\pi}_i$ is the marginal probability that channel i is available conditioned on previous observations and actions.

We can further show that this sufficient statistic $\tilde{\pi}$ takes only discrete values. It is equivalent to $\{[l_k, h_k]\}_{k=1}^{N}$ where $l_k \in \{0, 1\}$ is the last observed state of channel k and h_k is the time (in time slots) since this observation has been made. Namely, if $[l_k, h_k]$ is known for channel k, then the probability that channel k is available can be calculated as

$$[1-\tilde{\pi}_k, \tilde{\pi}_k] = [1-l_k, l_k](P_k)^{h_k}, \tag{7}$$

where $(P_k)^{h_k}$ is the h_k-step transition matrix. With the increase of h_k the convergence of $\tilde{\pi}_k$ to the stationary distribution $\tilde{\pi}_k^\infty$ is dominated by the second largest eigenvalue λ of P_k for any initial distribution $\tilde{\pi}_0$. Namely,

$$|\tilde{\pi}_k^\infty - (P_k)^{h_k}\tilde{\pi}_0| < D|\lambda|^{h_k} \tag{8}$$

for some positive constant D. For a 2-dimensional transition probability matrix $P_k = \{\{1-\alpha_k, \alpha_k\}, \{\beta_k, 1-\beta_k\}\}$, the second largest eigenvalue, also referred to as channel memory [16], is given as $\lambda = 1 - \alpha_k - \beta_k$. Therefore, if alternations between channel states are more often, *i.e.,* α_k and β_k are close to 0.5, the convergence to the stationary distribution is faster.

Assume now that if channel is not sensed for M consecutive slots, its distribution converges with sufficient accuracy to its stationary distribution[4]. In other words, if $h_k = M$, the last observation becomes irrelevant as it does not bear much information on the present information state of channel k. For notational convenience, we will set $l_k = 0$ if $h_k = M$. Therefore the state space of such a truncated problem for a certain channel k is

$$\mathcal{S}^k = \{(l,h) | l = 0,1;\; h = 0,1,\ldots,M\} \setminus \{(1,M)\}. \tag{9}$$

The composite state space of all channels is the Cartesian product of state spaces of all individual channels, i.e.

$$\mathcal{S} = \mathcal{S}^1 \times \mathcal{S}^2 \times \cdots \mathcal{S}^N. \tag{10}$$

The number of states in such composite state space is discussed in Section IV-C. An example of a possible state at a particular slot is shown in Fig. 3. In this example, it is assumed that there are $N = 7$ channels and that history tracking is truncated to $M = 4$ time slots. As illustrated in Fig. 3, we have

$$h_1 = 3,\; h_2 = 0,\; h_3 = h_5 = h_6 = h_7 = 4,\; h_4 = 1.$$

Note that there is no channel that is sensed 2 time slots prior to the current time slot in the example of Fig. 3. This is possible when channel 4 has been sensed in two successive time slots.

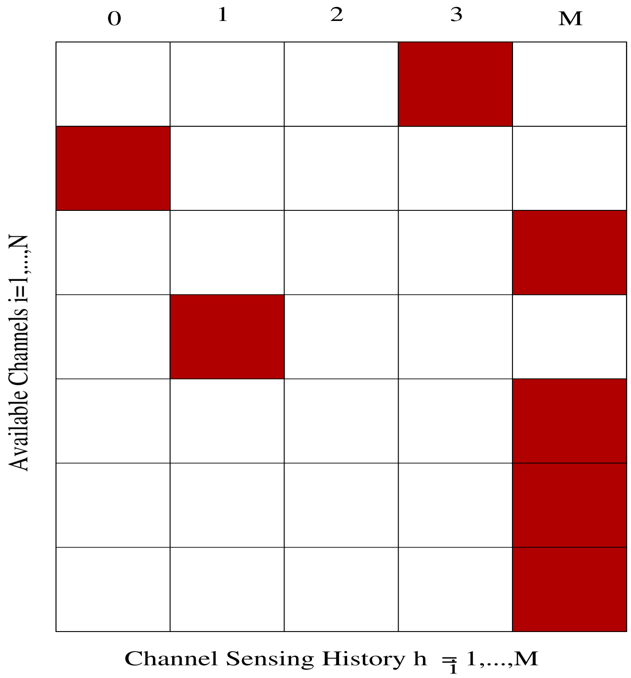

Fig. 3. Markovian channel model for the i-th channel.

B. Value Iteration

Let us first define the state transition operator $n(s,k,l)$ (given in (11)), where $s = \{(l_1,h_1),\ldots,(l_N,h_N)\} \in \mathcal{S}$ is the previous composite state (see (10)), $k \in 1,\ldots,N$ is the index of the channel being sensed, and $l = 0,1$ is the outcome of that observation. The operator $n(s,k,l)$ returns the next composite state.

Using the above notation, the Bellman's equation for the calculation of the optimal value function is then given in (12) for each $s \in \mathcal{S}$ of the form $s = \{(l_1,h_1),\ldots,(l_N,h_N)\}$.

In (12), $c(a,s)$ is the reward of applying action a in state $s = \{(l_1,h_1),\ldots,(l_N,h_N)\} \in \mathcal{S}$ and can be obtained as

$$c(a,s) = B_a \tilde{\pi}_{a|l_a,h_a}, \tag{13}$$

where B_a is the bandwidth of the channel a and $\tilde{\pi}_{a|l_a,h_a}$ is the (marginal) probability that channel a is available conditioned on the last measurement of l_a obtained h_a time slots ago. The latter can be computed from the transition matrix P_a as follows.

$$[1-\tilde{\pi}_{a|l_a,h_a}, \tilde{\pi}_{a|l_a,h_a}] = \begin{cases} [1-l_a, l_a](P_k)^{h_a} & \text{if} \quad h_a < M \\ [1-\tilde{\pi}_a^{\infty}, \tilde{\pi}_a^{\infty}] & \text{if} \quad h_a = M \end{cases}, \tag{14}$$

where $\tilde{\pi}_a^{\infty}$ denotes the probability that channel a is available when it reaches the stationary distribution.

C. Computational Complexity

We now analyze the computational complexity of obtaining the optimal policy under the truncated MDP formulation. Similar to the analysis for the POMDP formulation given in Section III, the complexity is determined by the size of the state space $\mathcal{S}$ given in (10).

1) Number of states: An upper bound on the number of states will be $(2M)^N$, which is exponential with N. However, consider the structure of the problem, we can show that the number of states is polynomial with N.

We are interested in cases where $N \gg M$. Since only one channel can be sensed in each slot, we have $h_i \neq h_j$ for $i \neq j$, unless $h_i = h_j = M$. Define

$$K = |\{i | h_i \neq M\}| \tag{15}$$

that is, K is the number of channels that we have sensed within the last $M-1$ slots. We then obtain the number of states as

$$|\mathcal{S}| = \sum_{K=1}^{M-1} \binom{N}{K} \binom{K}{1} \binom{M-2}{K-1} (K-1)! \times 2^K \tag{16}$$

From a computational complexity vantage point, we are primarily interested in the case when $N \gg M$, i.e. when there are much more channels compared to the history truncation parameter M. In this case, it can be seen that the first element in the product and the last element of the sum dominates the previous expression. Therefore, $|\mathcal{S}| \sim \mathcal{O}(N^{M-1})$.

2) Value Iteration Complexity: Note first that value iteration update (12) has to be calculated for all S states. Furthermore, in each update there are N possible choices for the action and there are only 2 summands within the sum with one multiplication. Therefore, the complexity of each iteration when only one channel is chosen for sensing is:

$$|\mathcal{S}| \times N \times 2 \tag{17}$$

For a horizon length of T, the total computational complexity C is given by

$$C \sim \mathcal{O}(N^M T),$$

which is polynomial with the number N of channels and linear with the horizon length T, in contrast to the exponential (with respect to T) complexity of the POMDP policies given in (5).

---

[4] For a given accuracy, we can obtain M based on the convergence rate of the underlying Markov process. See the footnote on Page 1.

$$
\begin{aligned}
n(s,k,l) \;=\; & \big\{(1+(l_1-1)I_{\{h_1<M\}}, \max(h_1+1,M)), \ldots, (1+(l_{k-1}-1)I_{\{h_{k-1}<M\}}, \max(h_{k-1}+1,M)), (l,0), \\
& (1+(l_{k+1}-1)I_{\{h_{k+1}<M\}}, \max(h_{k+1}+1,M)), \ldots, (1+(l_N-1)I_{\{h_N<M\}}, \max(h_N+1,M))\big\} \qquad (11)
\end{aligned}
$$

$$
V(s) = \max\left(c(j,s) + \sum_{l_j^{n+1}} p_1(l_j^{n+1}|l_j^n, h_j^n+1)V\left(n(s,j,l_j^{n+1})\right), j=1,\ldots,N\right) \qquad (12)
$$

To gain insight into the complexity reduction facilitated by the truncated MDP formulation of the spectrum access problem, let us consider the following example. Let the number of channels be $N = 6$, the history truncation parameter $M = 4$ and let the time horizon be $T = 20[slots]$. The POMDP complexity over the whole horizon is $2N^{T-1} = 3.8 \times 10^{21}$. At the same time the complexity of computing the policy via the truncated MDP formulation according to (17) is 24240 over the horizon of $T = 20$. The savings is over 15 orders of magnitude!

V. EXTENSIONS AND DISCUSSIONS

A. *Extensions to Sensing L out of N channels*

The value iteration (12) can be easily extended to the case of sensing L out of N channels[5]. In this case the action set consists of $\binom{N}{L}$ actions and it is necessary to change the state update equation given in (11). The details are omitted for brevity and to avoid repetitions.

However, as opposed to the case of $L = 1$, it is not easy to derive the exact expression for the number of states in the composite state of the truncated MDP model with $L > 1$. However, a simple upper bound can be derived as follows. Similarly to the case $L = 1$, let us define

$$K = |\{i | h_i \neq M\}| \qquad (18)$$

that is, K is the number of channels that we have sensed within the last $M - 1$ slots. Note that K is at least equal to L (since L channels have to be sensed in the current time slot) and at most $(M-2)L$ if L different channels are sensed in each of the $M - 2$ proceeding time slots. Therefore,

$$|\mathcal{S}| < \sum_{K=L}^{(M-2)L} \binom{N}{L} \binom{(N-L)(M-2)}{K-L} 2^K \qquad (19)$$

where the first binomial coefficient denotes the number of combinations to choose L out of N channels sensed in the current time slot. The second binomial coefficient denotes the number of combinations the remaining channel sensing histories can be chosen.

The complexity of the value iteration (in terms of the number of multiplications) in that case is as follows. In each iteration there are $\binom{N}{L}$ possible choices for the action and there are 2^L elements in the sum since we want to sum through all possible new channel realizations. Each element of the sum will comprise of L multiplications. Therefore, the complexity of each iteration when L channels are sensed is:

$$|\mathcal{S}| \times \binom{N}{L} \times 2^L \times L \qquad (20)$$

The total complexity has the same order: $C \sim \mathcal{O}(N^M T)$.

To illustrate the computational complexity needed for the calculation of the optimal L out of N sensing policy, let us consider the following example. Let the number of channels be $N = 6$, the history truncation parameter $M = 4$ and let the time horizon be $T = 20[slots]$.

Then, according to the equations (20) and (19), the number of multiplications for calculation of the optimal policy over the horizon of length T is upper bounded by 48240, 2036400, 2332800, 388800 for $L = 1, 2, 3, 4$, respectively. It can be seen that this these complexities are much smaller than the complexity of calculation of the optimal POMDP policy.

B. *Extensions to dependent channels*

Unfortunately, the truncated MDP model for OSA does not easily extend to the case of dependent channels. The reason is that state description of the channels using $[l_k, h_k], k = 1, \ldots, N$ does not in general provide a sufficient statistic for calculating the present information state. Specifically, if channels are correlated, sensing of one channel provides information for all other channels. One possible way out is to cluster together the channels that have similar statistical characterizations and consider them as one.

VI. NUMERICAL RESULTS

In this section we numerically explore the performance of the optimal and suboptimal opportunistic spectrum access strategies discussed above.

Experiment 1. (Influence of the Truncation Parameter M) In Fig. 4 we show the influence of the history truncation parameter M on the performance of spectrum access policies. The performance measure is the average throughput of the secondary users per time slot and we examine its dependency on the length of the horizon T and truncation history parameter M.

It is assumed that there are $N = 2$ identically and independently distributed channels with transition matrices $T_{1,2} = \{\{0.7, 0.3\}, \{0.2, 0.8\}\}$. The bandwidths of the channels are $B(1) = 5$ and $B(2) = 30$. The figure demonstrates that with the increase of parameter M, the average throughput of secondary users increases. Also, it is clear that the performance of the suboptimal history truncation scheme is very close to the performance of the optimal POMDP scheme. Furthermore, the performance of the OSA protocols developed under both the POMDP and the truncated MDP formulations improves over time, which results from the increasingly accurate information on the system state drawn from accumulating observations. This demonstrates the cognitive nature of the OSA protocols developed under the sequential decision-making frameworks of POMDP and MDP.

---

[5] To simplify notation, we consider $L_1 = L_2 = L$. Extension to $L_1 > L_2$ is straightforward.

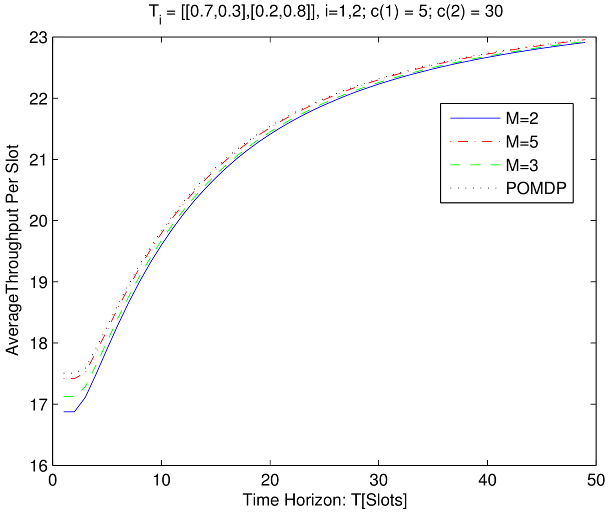

Fig. 4. Influence of truncation parameter M for fixed number of channels on the average throughput of the secondary users under the optimal policy.

Experiment 2. (Influence of the Number of Channels N) It is of interest to explore how does the throughput of secondary users behave with the increase of the number of utilized channels. To make a fair comparison, it is assumed that all channels have the same bandwidth $B(i) = 10, i = 1, \ldots, N$ and that all N channels are identically and independently distributed with transition probability matrix $T_i = \{\{0.7, 0.3\}, \{0.2, 0.8\}\}, i = 1, \ldots, N$.

From Fig. 5 it can be seen that increasing of the number of available channels increases the throughput of the secondary users. However, with the increase of the time horizon length, the gains caused by adding new channels are diminishing and the increase in per-slot throughput is saturating.

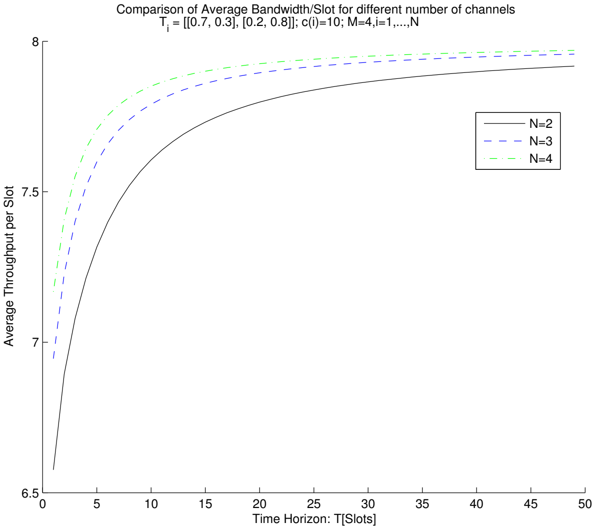

Fig. 5. Influence of the number of channels N for fixed M on the average throughput of the secondary users under the optimal policy.

VII. CONCLUSION

In this paper, we address distributed cognitive MAC protocols for opportunistic spectrum access. Our focus is the tradeoff between optimality and complexity. We show that the optimal design under the POMDP framework has a complexity that grows exponentially with the horizon length. To achieve a favorable tradeoff between optimality and complexity, we develop a truncated MDP formulation by exploiting the mixing time of the underlying Markovian model of spectrum usage. Analytical and numerical results demonstrate that the OSA protocol developed under the truncated MDP formulation has negligible performance loss yet a linear complexity with respect to the horizon length. Furthermore, by choosing an appropriate truncation parameter, we can determine the specific operating point on the tradeoff curve of performance vs. complexity.

REFERENCES

[1] M. McHenry, "Spectrum white space measurements," June, 2003. Presented to New America Foundation Broadband Forum; Measurements by Shared Spectrum Company, `http://www.newamerica.net/Download_Docs/pdfs/Doc_File_185_1.pdf`.

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

[3] FCC03-322, "Notice of Proposed Rule Making: Facilitating Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies and Authorization and Use of Software Defined Radios," December 2003.

[4] FCC04-186, "Notice of Proposed Rule Making: Unlicensed Operation in the TV Broadcast Bands," May 2004.

[5] A. Ghasemi and E. Sousa, "Collaborative Spectrum Sensing for Opportunistic Access in Fading Environments," in *Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2005.

[6] S. S. N, C. Cordeiro, and K. Challapali, "Spectrum Agile Radios: Utilization and Sensing Architectures," in *Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2005.

[7] A. Larcher, H. Sun, M. van der Shaar, and Z. Ding, "Decentralized Transmission Strategies for Delay-sensitive Applications over Spectrum Agile Networks ," in *Proc. of International Packet Video Workshop*, (Irvine, CA), Dec. 2004.

[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] M. Steenstrup, "Opportunistic Use of Radio-Frequency Spectrum: A Network Perspective," in *Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2005.

[10] 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.

[11] 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.

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

[13] Y. Chen, Q. Zhao, and A. Swami, "Joint Design and Separation Principle for Opportunistic Spectrum Access in the Presence of Sensing Error," submitted to *IEEE Transactions on Information Theory*, Feb., 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] M.Zorzi and R.Rao, "Error control and energy consumption in communications for nomadic computing," *IEEE Transactions on Computers*, vol. 46, pp. 279–289, March 1997.